



**UNIVERSIDADE FEDERAL DE SANTA CATARINA  
PROGRAMA DE PÓS-GRADUAÇÃO EM ENGENHARIA DE  
PRODUÇÃO**

Rafael Tezza

**MODELAGEM MULTIDIMENSIONAL PARA MENSURAR  
QUALIDADE EM WEBSITE DE E-COMMERCE UTILIZANDO A  
TEORIA DA RESPOSTA AO ITEM**

Tese submetida ao Programa de Pós-Graduação em Engenharia de Produção da Universidade Federal de Santa Catarina para a obtenção do Grau de Doutor em Engenharia de Produção.

Orientador: Prof. Dr. Antonio Cezar Bornia

Florianópolis  
2012

Ficha de identificação da obra elaborada pelo autor,  
através do Programa de Geração Automática da Biblioteca Universitária da UFSC.

Tezza, Rafael

Modelagem multidimensional para mensurar qualidade em  
website de e-commerce utilizando a teoria da resposta ao  
item [tese] / Rafael Tezza ; orientador, Antonio Cezar  
Bornia - Florianópolis, SC, 2012.

182 p. ; 21cm

Tese (doutorado) - Universidade Federal de Santa  
Catarina, Centro Tecnológico. Programa de Pós-Graduação em  
Engenharia de Produção.

Inclui referências

1. Engenharia de Produção. 2. qualidade em websites. 3.  
e-commerce. 4. dimensionalidade. 5. teoria da resposta ao  
item multidimensional. I. Bornia, Antonio Cezar. II.  
Universidade Federal de Santa Catarina. Programa de Pós-  
Graduação em Engenharia de Produção. III. Título.

Rafael Tezza

**MODELAGEM MULTIDIMENSIONAL PARA MENSURAR  
QUALIDADE EM WEBSITES DE E-COMMERCE UTILIZANDO A  
TEORIA DA RESPOSTA AO ITEM**

Esta Tese foi julgada adequada para obtenção do Título de “Doutor” e aprovada em sua forma final pelo Programa de Pós-Graduação em Engenharia de Produção.

Florianópolis, 17 de dezembro de 2012.

---

Prof. Antonio Cezar Bornia, Dr.  
Coordenador do Curso

**Banca Examinadora:**

---

Prof. Antonio Cezar Bornia, Dr.  
Orientador  
EPS/UFSC

---

Prof. Dalton Francisco de Andrade, PhD.  
INE/UFSC

---

Prof. Júlio da Silva Dias, Dr.  
Examinador Externo  
UDESC

---

Profa. Lizandra Garcia Lupi Vergara, Dra.  
EPS/UFSC

---

Prof. Pedro Alberto Barbeta, Dr.  
INE/UFSC

---

Profa. Silvana Ligia Vincenzi Bortolotti, Dra.  
Examinadora Externa  
UTFPR

## AGRADECIMENTOS

Ao meu orientador, prof. Antonio Cezar Bornia pela confiança incentivos e ensinamentos. Ao prof. Dalton Francisco de Andrade por compartilhar seus conhecimentos e sua paixão pela Teoria da Resposta ao Item. Ao prof. Pedro Alberto Barbeta, pelas valiosas discussões sobre dimensionalidade e softwares estatísticos. Às professoras Lizandra Garcia Lupi Vergara e Silvana Ligia Vincenzi Bortolotti e ao prof. Júlio da Silva Dias por terem se disponibilizado a avaliar este trabalho nesta reta final de 2012.

Aos colegas da ESAG/UDESC que me ajudaram direta ou indiretamente, e sempre tiveram paciência para ouvir minhas desculpas de falta de tempo, dada a jornada dupla de ensinar e escrever este trabalho.

Às três mulheres de minha vida, minha companheira Júlia Ferreira, minha mãe Maria Terezinha Tezza e minha avó Maria Benta Mafra Tezza que, pacientemente me permitiram viver na ponta dos pés durante a conclusão deste trabalho.

Aos amigos Rafael Marques de Albuquerque, Michele Afonso e Júlio da Silva Dias pelas valiosas sugestões e revisões e ao “seu Búrigo” pela fonte de energias positivas. Aos colegas do curso, e aos amigos Andréa Cristina Trierweiler, Blênio Severo Peixe, Débora Spenassato, Fernando Moreira Júnior, Luciano Rath, Rosemeri Maria de Souza e a todos os professores e funcionários do Programa de Pós-Graduação em Engenharia de Produção da UFSC.

À CAPES e ao CNPq pelo financiamento nos primeiros dois anos do curso e a todos que ajudaram ou torceram para que este trabalho se tornasse real.

## RESUMO

O presente trabalho tem por objetivo construir uma modelagem multidimensional para medir a qualidade de websites comerciais com uso da Teoria da Resposta ao Item (TRI). Como qualidade de um website, entende-se suas características técnicas (usabilidade/navegabilidade, apresentação da informação e interatividade) e não técnicas (design, estética, apelo visual, confiabilidade, hedonismo, imagem) configurando, teoricamente, um contexto multidimensional. As suposições iniciais das dimensões e elaboração dos itens foi baseada em uma análise bibliográfica sobre o tema qualidade em websites de e-commerce. Um conjunto de 75 itens foi elaborado e submetido a uma amostra de 441 websites de e-commerce dos mais variados setores. O tratamento e análise dos dados foi feita utilizando-se a Teoria da Resposta ao Item. Nesta etapa, são discutidas questões referentes à dimensionalidade e a escolha do modelo mais adequado. Por fim, um modelo multidimensional com quatro dimensões é ajustado compreendendo as dimensões de orientação do usuário e navegabilidade; acessibilidade e confiabilidade do sistema; interatividade; e apresentação da informação.

**Palavras-chave:** qualidade em websites, e-commerce, dimensionalidade, teoria da resposta ao item multidimensional.

## ABSTRACT

This paper aims to build a multidimensional modeling to measure the quality of commercial websites using the Item Response Theory (IRT). As a quality website means its technical characteristics (usability / navigation, presentation of information and interactivity) and non-technical (design, aesthetics, visual appeal, reliability, hedonism, image) setting, theoretically, a multidimensional context. The initial assumptions of the dimensions and design of the items was based on a literature review on the topic of quality in e-commerce websites. A set of 75 items was developed and submitted to a sample of 441 e-commerce websites in several sectors. Treatment and data analysis was done using the Item Response Theory. At this stage, we discuss issues of dimensionality and choose the most suitable model. Finally, a multidimensional model is fitted with four dimensions comprising the dimensions of user orientation and navigability, accessibility and reliability of the system, interactivity, and presentation of information.

**Keywords:** quality websites, e-commerce, dimensionality, item response theory multidimensional.

## Lista de Figuras

Figura 1 - Distribuição de frequência das abordagens levantadas no Quadro 1 quanto às dimensões mais citadas.....	36
Figura 2 - Principais estágios de um survey .....	41
Figura 3 - Curva Característica do Item para modelo logístico de um parâmetro....	46
Figura 4 - Curva Característica do Item para modelo logístico de dois parâmetros.	47
Figura 5 - Distinção entre a multidimensionalidade dentro de cada item e entre itens.....	57
Figura 6 - Quatro possíveis estruturas de traços latentes, Reise et al. (2007). ....	65
Figura 7 - Estrutura do modelo de segunda ordem da TRIM, Rijmen, 2010. ....	67
Figura 8 - Sistemática metodológica. Adaptada de Wilson (2005).....	71
Figura 9 - Método para o levantamento bibliográfico utilizado. Adaptado de Villas et al. (2008).....	72
Figura 10 - Valores próprios da matriz de correlação tetracórica considerando 10 dimensões para o fator latente.....	89
Figura 11 - Resultado da análise paralela para os 75 itens.....	90
Figura 12: Valores próprios da matriz de correlação tetracórica considerando os 44 itens.....	100
Figura 13 - Análise paralela dos 44 itens restantes. ....	101
Figura 14 - Representação gráfica da relação de cada um dos 44 itens restantes com as quatro dimensões caracterizadas. A linha contínua representa a ligação de maior intensidade e as pontilhadas de menor. ....	134

## **Lista de Quadros**

Quadro 1- Classificação dos estudos mais relevantes sobre qualidade em e-service, de 2000 à 2012.....	30
Quadro 2 – Descrição dos itens formulados e referidas referências diretas.....	74
Quadro 3 – Composição da amostra segundo o tipo de produto comercializado pelo website.....	80
Quadro 4 – Descrição dos 31 itens eliminados.....	95
Quadro 5 – Classificação dos itens conforme a dimensão, segundo a análise fatorial .....	110
Quadro 6 - Classificação dos itens conforme a dimensão, segundo o modelo multidimensional da teoria da resposta ao item.....	135



## Lista de Tabelas

Tabela 1 - Valores próprios da matriz de correlação tetracórica.....	88
Tabela 2 - Comparação dos modelos de um, dois, três e quatro dimensões. ....	91
Tabela 3 - Índice de AIC e BIC para os modelos de um, dois, três e quatro dimensões. ....	92
Tabela 4 - Correlação dos fatores assumindo três dimensões aos 75 itens. ....	93
Tabela 5 - Análise fatorial de informação plena, com rotação ortogonal varimax...93	
Tabela 6 - Valores próprios da matriz de correlação tetracórica dos primeiros 10 autovalores do conjunto de 44 itens. ....	100
Tabela 7 - Seleção do melhor modelo com base nos critérios de informação AIC e BIC.....	101
Tabela 8 - Correlação linear dos fatores assumindo três dimensões dos 44 itens..	102
Tabela 9 - Análise fatorial utilizando o método da informação plena, assumindo rotação oblíqua promax e ortogonal varimax aos 44 itens em três dimensões. ....	102
Tabela 10: Correlação dos fatores assumindo quatro dimensões dos 44 itens.....	104
Tabela 11 - Análise fatorial utilizando o método da informação plena, assumindo rotação ortogonal varimax aos 44 itens em quatro dimensões. ....	105
Tabela 12 - Correlação dos fatores assumindo 5 dimensões dos 44 itens.....	111
Tabela 13 - Análise fatorial utilizando o método da informação plena, assumindo rotação ortogonal varimax e oblíqua promax aos 44 itens em cinco dimensões. ....	111
Tabela 14 - Estimação dos parâmetros de dificuldade e discriminação assumindo o modelo unidimensional de dois parâmetros para os 44 itens. ....	114
Tabela 15 - Estimação dos parâmetros de dificuldade e discriminação assumindo o modelo unidimensional de dois parâmetros ModU1.....	115
Tabela 16 - Estimação dos parâmetros de dificuldade e discriminação assumindo o modelo unidimensional de dois parâmetros ModU2.....	116
Tabela 17 - Estimação dos parâmetros de dificuldade e discriminação assumindo o modelo unidimensional de dois parâmetros ModU3.....	117
Tabela 18 - Estimação dos parâmetros de dificuldade e discriminação assumindo o modelo unidimensional de dois parâmetros ModU4.....	118

Tabela 19 - Estimação dos parâmetros de dificuldade e discriminação assumindo o modelo unidimensional de dois parâmetros ModU24.....	119
Tabela 20 - Estimação dos parâmetros de dificuldade e discriminação assumindo o modelo unidimensional de dois parâmetros ModU13.....	120
Tabela 21 - Estimação dos parâmetros de discriminação para cada dimensão e parâmetro de discriminação multidimensional (MDISC), parâmetro de dificuldade escalar (d) o parâmetro de dificuldade multidimensional (MDIFF) e seus respectivos erros padrões (se) para os 44 itens, assumindo o modelo multidimensional de dois parâmetros compensatório.....	122
Tabela 22 - Reestimação dos parâmetros de discriminação para cada dimensão e parâmetro de discriminação multidimensional (MDISC), parâmetro de dificuldade escalar (d) o parâmetro de dificuldade multidimensional (MDIFF) e seus respectivos erros padrões (se) para os 42 itens, assumindo o modelo multidimensional de dois parâmetros compensatório.....	124
Tabela 23 - Reestimação dos parâmetros de discriminação para cada dimensão e parâmetro de discriminação multidimensional (MDISC), parâmetro de dificuldade escalar (d) o parâmetro de dificuldade multidimensional (MDIFF) e seus respectivos erros padrões (se) para os 40 itens, assumindo o modelo multidimensional de dois parâmetros compensatório.....	126
Tabela 24 - Ranqueamento dos itens segundo os parâmetros de discriminação (MDISC) e dificuldade (MDIFF) de multidimensional. ....	136
Tabela 25 - Estimação do grau de qualidade multidimensional dos 4 primeiros webistes da amostra com base no modelo multidimensional de dois parâmetros compensatório. ....	137
Tabela 26 - Probabilidade de resposta positiva dos 40 itens para os 4 primeiros websites analisados com base nos seus respectivos graus de qualidade conforme estimado no modelo multidimensional compensatório de quatro dimensões. ....	138
Tabela 27 - Estimação dos parâmetros de discriminação (a) para cada dimensão e na dimensão geral e parâmetro de dificuldade escalar (d) assumindo o modelo bi-fatorial. ....	139
Tabela 28 - Comparação entre os modelos unidimensional, multidimensional de dois parâmetros compensatório e do modelo bi-fatorial com base nos critérios de informação AIC e BIC, RMSEA e M2. ....	142

## **Lista de siglas e abreviaturas**

AF – Análise Fatorial

AIC - *Akaike Information Criterion*

Anova – Análise de variância

BIC - *Bayesian Information Criterion*

CCI – Curva Característica do Item

*e-commerce – electronic commerce*

*e-service – electronic service*

FI – Full information

MDIFF – Parâmetro de dificuldade multidimensional

MDISC – Parâmetro de discriminação multidimensional

MEE – Modelagem de equação estruturais

ML1 – Modelo logístico de um parâmetro

ML2 – Modelo logístico de dois parâmetros

ModU1 – Modelo unidimensional considerando apenas os itens da dimensão um da análise fatorial com quatro dimensões

ModU2 – Modelo unidimensional considerando apenas os itens da dimensão dois da análise fatorial com quatro dimensões

ModU3 – Modelo unidimensional considerando apenas os itens da dimensão três da análise fatorial com quatro dimensões

ModU4 – Modelo unidimensional considerando apenas os itens da dimensão quatro da análise fatorial com quatro dimensões

MRCMLM – *Multidimensional Random Coefficients Multinomial Logit Model*

PISA - *Programme for International Student Assessment*

RMSEA - Raiz do erro médio quadrático de aproximação

TCT – Teoria clássica dos testes

TRI – Teoria da resposta ao item

TRIM – Teoria da resposta ao item multidimensional

W3C – *World Wide Web Consortium*

WCAG - *Web Content Accessibility Guidelines*

## SUMÁRIO

<b>1 INTRODUÇÃO.....</b>	<b>15</b>
1.1 CONTEXTUALIZAÇÃO.....	15
1.2 PROBLEMA DA PESQUISA.....	16
1.3 OBJETIVOS .....	18
1.3.1 Objetivo Geral.....	18
1.3.2 Objetivos Específicos.....	18
1.4 JUSTIFICATIVA.....	19
1.4.1 Relevância .....	19
1.4.2 Ineditismo.....	21
1.5 LIMITES DO TRABALHO .....	22
1.6 ESTRUTURA DO TRABALHO .....	22
<b>2 REFERENCIAL TEÓRICO .....</b>	<b>24</b>
2.1 QUALIDADE NA WEB.....	24
2.2 CRIAÇÃO DE ESCALAS .....	37
2.2.1 Tipologias de escalas.....	38
2.2.2 Vantagens e desvantagens das escalas .....	39
2.2.3 Elaboração de um conjunto de itens .....	40
2.3 TEORIA DA RESPOSTA AO ITEM.....	42
2.3.1 Modelos unidimensionais da TRI.....	45
2.3.2 Dimensionalidade e Teoria da Resposta ao Item.....	49
2.3.3 Modelos Multidimensionais da TRI.....	54
2.4 SÍNTESE DO CAPÍTULO.....	67
<b>3. MÉTODO .....</b>	<b>69</b>
3.1 CARACTERIZAÇÃO DA PESQUISA.....	69
3.2 ELABORAÇÃO DOS ITENS .....	70
3.3 COLETA DOS DADOS .....	79
3.4 METODOLOGIA ESTATÍSTICA .....	82
3.4.1 - Avaliação do número de dimensões.....	83
3.4.2 - Avaliação da qualidade dos itens .....	84
3.4.3 - Validação das dimensões .....	84
3.4.4 - Verificação da adequação do modelo unidimensional e unidimensional múltiplo.....	84
3.4.5 - Verificação da adequação do modelo de multidimensional compensatório .....	85
3.4.6 - Verificação da adequação do modelo bi-fatorial.....	85
3.5 SÍNTESE DO CAPÍTULO.....	86

<b>4. RESULTADOS E DISCUSSÕES .....</b>	<b>88</b>
4.1 AVALIAÇÃO DO NÚMERO DE DIMENSÕES.....	88
4.2 AVALIAÇÃO DA QUALIDADE DO CONJUNTO DE ITENS INICIAL .....	92
4.3 VALIDAÇÃO DAS DIMENSÕES.....	99
4.4 VERIFICAÇÃO DA ADEQUAÇÃO DO MODELO UNIDIMENSIONAL E UNIDIMENSIONAL MÚLTIPLO .....	113
4.5 - VERIFICAÇÃO DA ADEQUAÇÃO DO MODELO MULTIDIMENSIONAL COMPENSATÓRIO .....	121
4.6 - VERIFICAÇÃO DA ADEQUAÇÃO DO MODELO BI-FATORIAL .....	139
4.7 SÍNTESE DO CAPÍTULO .....	142
<b>5. CONCLUSÃO .....</b>	<b>145</b>
5.1 CONCLUSÕES .....	145
5.2 RECOMENDAÇÕES .....	147
<b>REFERÊNCIAS .....</b>	<b>149</b>
<b>APÊNDICE.....</b>	<b>174</b>

# 1 INTRODUÇÃO

## 1.1 CONTEXTUALIZAÇÃO

O mercado de comércio eletrônico (e-commerce) vem crescendo exponencialmente desde o início da década passada em todo o mundo, aumentando assim seu poder de difusão e competitividade (KIM, et al., 2012). O Brasil é o quinto país com maior número de usuários de internet (Internet World Stats, 2011), o que potencialmente representa um dos maiores mercados de e-commerce global.

Segundo e-bit (2012), no Brasil, o e-commerce cresceu 21% no primeiro semestre de 2012 em relação ao mesmo período de 2011, atingindo faturamento de R\$ 10,2 bilhões e valor médio de compras no valor de R\$ 346,00. Para o segundo semestre de 2012, a expectativa se mantém nesse patamar e o faturamento deve chegar a R\$ 12,2 bilhões.

Do ponto de vista empresarial e mercadológico, as empresas de e-commerce representam uma cadeia de processos que se inicia com a visita do cliente ao website da empresa, passando pela interação deste no website, efetivação da compra, processo logístico envolvido na entrega do produto, no caso de produtos tangíveis, e eventuais serviços de pós-venda. Cada um destes processos possui sua particular importância. O website no qual os produtos são visualizados é equivalente à loja física onde é mostrada a mercadoria e a empresa comunica sua imagem para o consumidor por meio de sua interface gráfica. Segundo Downes (2007), o website é frequentemente o primeiro ponto de contato entre uma empresa que usa e-commerce e seus consumidores. Desta forma, manter um site de compras atraente e amigável é essencial para o sucesso de qualquer empresa na internet (McKINNEY, 2004). Além disso, segundo Deng e Poole (2012), como o uso da internet continua a crescer, as empresas reconhecem cada vez mais a necessidade competitiva de ter uma presença na web e se engajar neste mercado. Isso tem estimulado um crescimento maciço de sites comerciais e intensificado a concorrência por clientes on-line, gerando uma proliferação de sites fornecendo funcionalidades similares e produtos, serviços ou informações idênticos, permitindo que os clientes possam mudar de um site para outro sem esforço. Assim, para atrair e reter clientes por meio da internet, uma organização precisa fazer seu website o mais simples e atraente para o uso possível; caso contrário, os clientes potenciais podem simplesmente ir para outro website (TAYLOR e ENGLAND, 2006).

Segundo Alba et al. (1997), as peculiaridades do comércio eletrônico em relação às lojas de varejo tradicional são: (1) a janela de visão é restrita, ao invés de andar em um grande espaço físico da loja, o ambiente de compras é uma pequena tela, (2) distância e tempo são comprimidas, e (3) os consumidores têm mais controle sobre as informações que procuram e os sites que eles visitam.

Neste contexto, considerando estas peculiaridades e o grande crescimento das empresas de comércio eletrônico, medidas de avaliação da qualidade destes serviços têm se tornado cada vez mais importante (DING et al. 2011).

Para Aladwani e Palvia (2002), a construção de medidas de qualidade no contexto de websites é uma tarefa desafiadora, pois qualidade em websites é um conceito complexo e de natureza multidimensional. Segundo Loiacono et al. (2002) avaliação de qualidade em websites tem sido abordada por três ângulos principais: máquina (BAUER; SCHARL, 2000, TREIBLMAIE, 2007), avaliação de especialistas (SMITH, 2001; BÜYÜKÖZKAN et al., 2007) e percepção do consumidor (SCHUBERT; DETTLING, 2002; HUNG; McQUEEN, 2004). Os três ângulos envolvem características técnicas tais como velocidade de download (PALMER, 2002, GALLETTA, et al., 2004), segurança (FLAVIÁN; GUINALÍU, 2006; BRAZ et al., 2007), usabilidade (BANGOR et al., 2008; FANG et al., 2007), qualidade do conteúdo (AGARWAL; VENKATESH, 2002; GOLDBERG; ALLEN, 2008) e características não técnicas tais como estética (PANDIR; KNIGHT, 2006; LINDGAARD et al., 2006) confiabilidade (SEFFAH, et al., 2006; FLAVIÁN et al., 2006) e apelo visual (KULVIWAT et al., 2007; LINDGAARD, 2007).

A complexidade argumentada por Aladwani e Palvia (2002) pode ser a responsável pelo fato de que, apesar da óbvia importância do tema, a conceitualização e a medição de qualidade em websites estão ainda numa fase inicial de desenvolvimento (CRISTOBAL et al., 2007; FASSNACHT e KOESE, 2006; LADHARI, 2010). Além disso, segundo Parasuraman et al. (2005) e Ding et al. (2011), os estudos neste campo são um pouco limitados e díspares.

## 1.2 PROBLEMA DA PESQUISA

Websites comerciais estão se tornando cada vez mais complexos, pois o número de funcionalidades oferecidas aos consumidores está aumentando constantemente, a fim de melhorar tanto o processo de coleta



de informações quanto a experiência de compra online (ETHIER et al., 2008). Com estes novos desenvolvimentos, gestores e desenvolvedores de websites estão sobrecarregados de abordagens com recomendações sobre como criar e manter um website atraente e de qualidade.

Pesquisadores de várias áreas desenvolveram e testaram instrumentos para medir a qualidade de website (por exemplo, ALADWANI; PALVIA 2002; LOIACONO 2002; VAN DER MERWE; BEKKER, 2003). Nestes estudos, um grande número de itens foi gerado para medir a qualidade de websites e estes foram testados em uma variedade de sites comerciais ou não comerciais por meio de avaliação de usuários experientes, novatos, estudantes, desenvolvedores, entre outros, e os achados têm indicado que a qualidade em websites representa um construto multidimensional (KIM; STOEL, 2004). Entretanto, poucos estudos têm explorado a hierarquia das dimensões destes elementos, de modo que pouco se sabe sobre as relações entre as diversas escalas ou a construção global de uma escala para medir qualidade do website.

Esta lacuna se dá pelo fato de que a maioria das abordagens apresenta algumas deficiências. Primeiro, alguns conjuntos de propostas de diretrizes são frequentemente contraditórios (FROHLICH, 2004; HASSENZAHL, 2004; KUROSU; KASHIMURA, 1995; MONK, LELOS, 2007; OVERBEEKE; WENSVEEN, 2004; KIM; STOEL, 2004), por exemplo, Hassenzahl (2004) não encontraram nenhuma relação entre juízos de qualidade estética e usabilidade enquanto Tractinsky et al. (2000) argumentam o contrário. Além disso, algumas são baseadas em fundamentos puramente teóricos (GEISLER et al., 2001; LIU; ARNETT, 2000; SONG; ZAHEDI, 2001; ZHANG; VON DRAN, 2000). Em segundo lugar, de acordo com Ethier et al. (2008), as recomendações de design para websites geralmente ignoram fatores relacionados à estética, apelo visual, emoções e hedonismo, apesar dos importantes resultados empíricos sobre o impacto crucial destes fatores no comportamento de compra, tanto na esfera tradicional quanto no ambiente online (BAGOZZI et al., 1999; KALBACH, 2006; TURLEY; MILLIMAN, 2000). Terceiro, de acordo com Hamilton (2006), a maioria dos modelos de análise de qualidade de websites têm falhado por tratar questões gerais de qualidade de websites como modelos lineares e unidimensionais. Além disso, este autor argumenta que os poucos estudos tratando da qualidade em websites não empregam procedimentos rigorosos e sistemáticos para desenvolver e validar um instrumento. Uma alternativa para resolver estes problemas seria o desenvolvimento de uma sistemática ou modelagem multidimensional considerando, da forma mais abrangente possível, os fatores que influenciam diretamente a qualidade de um website e elaborar uma medida padronizada. Neste sentido, a Teoria da

Resposta ao Item Multidimensional surge como uma alternativa para o desenvolvimento desta sistemática, uma vez que está associada a uma estrutura matemática e metodológica robusta, capaz de criar uma escala multidimensional onde o foco não é apenas a redução de fatores como na análise fatorial, sendo possível comparar item e respondente.

Partindo desta problemática, o presente estudo levanta a seguinte questão:

É possível desenvolver uma modelagem multidimensional robusta para mensurar qualidade em websites que concentre os achados empíricos e teóricos da literatura em uma escala padronizada capaz de posicionar itens e websites no mesmo contínuo?

### 1.3 OBJETIVOS

#### 1.3.1 Objetivo Geral

Desenvolver uma modelagem multidimensional utilizando a Teoria da Resposta ao Item para mensurar qualidade em websites comerciais.

#### 1.3.2 Objetivos Específicos

- Definir o construto qualidade em websites comerciais com base na literatura acadêmica;
- Identificar as dimensões mais relevantes em qualidade em websites comerciais;
- Desenvolver um conjunto de itens para avaliar qualidade em websites comerciais;
- Analisar a dimensionalidade no contexto de avaliação da qualidade em websites comerciais;
- Verificar a adequação dos modelos uni e multidimensionais da teoria da resposta ao item no contexto de avaliação de qualidade em websites comerciais;

## 1.4 JUSTIFICATIVA

### 1.4.1 Relevância

Pelo fato de que os consumidores on-line podem facilmente mudar para outros sites de comércio eletrônico, sua lealdade online geralmente não é elevada. Como atrair a atenção dos consumidores online e mantê-los no site tornou-se um grande desafio para as empresas de comércio eletrônico (XU; LIU, 2010). E este poder de atratividade está ligado diretamente à qualidade da interface.

Qualidade em websites é uma questão complexa e de múltiplas facetas que deve ser discutida por múltiplos ângulos o que vai além da simples facilidade de uso da interface (questões técnicas), contando com questões ligadas a fatores estéticos, de confiabilidade, interatividade, entre outros, que constituem questões não técnicas. Segundo Nielsen (2007), questões técnicas como a usabilidade continuam sendo condições necessárias, mas não condições suficientes em todos os contextos para o julgamento positivo da qualidade de um website. Liu et al. (2009) argumentam que existem muito poucos estudos sobre a qualidade em websites. Além disso, os estudos existentes argumentam que a qualidade em websites possui um construto multidimensional, mas não existe um consenso de quais dimensões são mais relevantes (ZEITHAML et al, 2002; PARASURAMAN et al, 2005; FASSNACHT; KOESE, 2006).

Embora existam na literatura várias pesquisas sobre comportamento do usuário na web, usabilidade, estética, interatividade, confiabilidade, qualidade da informação e personalização, estes levantamentos, assim como as avaliações de qualidade de websites, têm sido, de certa forma, fragmentados (MONSUWE et al. 2004; MBIPOM e HARPER, 2009) e utilizado tratamentos estatísticos tradicionais pouco robustos. Segundo Chiou et al. (2010), do ponto de vista geral, estas abordagens são bastante úteis, avaliando a atitude do usuário para com o site e pode ser considerada como uma avaliação externa do websites. No entanto, sob o ponto de vista estratégico, pouca atenção tem sido dada à avaliação da coerência entre estes achados e a realidade prática de uma organização de e-commerce.

Desta forma, torna-se relevante, tanto do ponto de vista teórico quanto prático, o desenvolvimento de uma sistemática ou modelagem que compile estas pesquisas e realize um tratamento estatístico robusto. Para isto, uma alternativa matematicamente viável dentro do contexto de avaliação de websites, segundo Schmettow e Vietze (2008) e Tezza et al.

(2011), é a Teoria da Resposta ao Item, a qual permite a mensuração de características difíceis de serem medidas diretamente (traços latentes) com base em um conjunto de itens e na criação de uma escala interpretável.

Do ponto de vista da relevância teórica, o presente estudo evidencia a discussão conjunta de questões referentes à conceituação e mensuração de qualidade de websites, sua dimensionalidade e o ajuste de uma modelagem estatística adequada. Para isso, é apresentada uma sistemática para desenvolvimento dos itens e algumas abordagens para tratamento de construtos multidimensionais de onde surge a modelagem teórica proposta.

Do ponto de vista da relevância prática, destaca-se o desenvolvimento de uma escala padronizada com interpretação direta na qual se possa verificar a relação entre os componentes de qualidade (itens), dimensões e o desempenho do website em um mesmo *continuum*. Com isto, desenvolvedores de websites comerciais podem verificar quais requisitos são mais relevantes e qual o impacto na alteração destes.

A escolha da utilização da teoria da resposta ao item (TRI) como instrumento para desenvolvimento desta modelagem para medir qualidade em websites se deu principalmente pelas suas características metodológicas. Ou seja, é capaz de desenvolver uma escala para mensurar traços latentes, como por exemplo, a qualidade de websites, tendo como base um conjunto de itens, os quais também podem ser posicionados na mesma escala do traço latente. Além disso, a TRI é capaz de analisar a dimensionalidade e a discriminação de um conjunto de itens, o que, segundo Bartolucci et al (2012) não é possível em nenhuma outra abordagem estatística conhecida, como por exemplo a teoria clássica dos testes, a análise fatorial (AF) e a modelagem de equações estruturais (MEE). No caso da teoria clássica dos testes, também utilizada para medir traços latentes, o fato desta ser sustentada basicamente pela relação entre escores observados e escores verdadeiros em um teste, concentra suas conclusões no teste e torna o instrumento dependente da amostra, o que não ocorre na teoria da resposta ao item, onde o foco não é o teste como um todo e sim o item e a escala gerada é independente da amostra. Quando se trata de construtos multidimensionais, a técnica de AF é uma das técnicas mais comuns, entretanto, como a teoria clássica dos testes, esta técnica não tem foco no respondente e tão pouco no item, servindo mais como uma forma de redução de dados ou uma alternativa para identificação de clusters. Segundo Mendes (2006), sob o ponto de vista analítico a teoria da resposta ao item possui algumas vantagens em relação aos modelos de equações estruturais, principalmente no que tange à análise individual e comparativa dos itens do conjunto.

Sendo assim, sumariamente, a TRI, tanto em sua modelagem simples, unidimensional, como a mais complexa, multidimensional, apresenta-se como uma ferramenta capaz de medir traços latentes, evidenciando as relações entre respondentes e itens, de forma a posicioná-los em uma mesma escala comparável e válida para situações mais amplas por não depender diretamente do grupo de respondentes.

Num segundo momento, a escolha de teoria da resposta ao item, particularmente em sua estrutura multidimensional apresenta-se como uma alternativa inédita no contexto de sistemas, mais especificamente em websites. Segundo Allen e Wilson (2006), apesar da evolução na literatura metodológica, a aplicação de modelos multidimensionais da TRI em áreas diferentes da educacional tem sido limitada. Isto se dá provavelmente devido (i) aos problemas de complexidade nos cálculos estatísticos que estão envolvidos desses modelos e (ii) a dificuldade associada à interpretação multidimensional destes modelos.

#### 1.4.2 Ineditismo

Segundo Rossiter (2009) se as pesquisas em e-commerce estão se tornando uma “verdadeira ciência”, estas devem ser baseadas em medidas validadas e confiáveis dentro de construtos chave. Esta preocupação de Rossiter vai ao encontro da necessidade de explorar metodologias e maneiras alternativas de consolidar as pesquisas nesta área, de forma a sistematizar o conhecimento evidenciando os fatores envolvidos de forma mensurável e, conseqüentemente, com poder de reprodutibilidade.

No entanto, neste contexto, raramente são encontradas sistematizações ou modelagens que considerem além das preferências do usuário, que são de certa forma medidas subjetivas. Desta forma, a presente pesquisa busca suprir esta lacuna, propondo a criação de uma sistemática objetiva que considera as pesquisas realizadas na última década baseadas em preferências dos usuários, transformando-as em uma modelagem objetiva, o que representa uma visão inexplorada até então na literatura. Além disso, esta tradução e consolidação da literatura são discutidas e modeladas sob a perspectiva da teoria da resposta ao item multidimensional até então inédita na literatura de qualidade em websites.

A utilização dos modelos da teoria da resposta ao item multidimensional (TRIM) para lidar com problemas de medição em avaliação educacional em larga escala tem sido realizada desde o início dos anos 1990 (ACKERMAN, 1992; CAMILLI, 1992; EMBRETSON, 1991; GLAS, 1992; OSHIMA, MILLER, 1992; RECKASE; McKINLEY, 1991).

No entanto, segundo Adams, et al. (1997), Hartig, Höhler (2008), Rauch e Hartig (2010), a aplicação de modelos em teste prático fora da área educacional é relativamente rara. Na avaliação de sistemas, mais especificamente sistemas web, as aplicações da TRI são muito raras, Schmettow e Vietze (2008) propõem a aplicação do modelo de Rasch (RASCH, 1960) para o processo de inspeção de usabilidade, Tezza et al. (2011) aplicam o modelo logístico de dois parâmetros unidimensional (BIRNBAUM, 1968) para mensurar usabilidade em websites comerciais. Não é encontrada na literatura aplicação de modelos mais complexos da TRI que considerem a multidimensionalidade na avaliação de websites comerciais.

## 1.5 LIMITES DO TRABALHO

- A elaboração dos itens, e identificação *a priori* das dimensões é feita, como a grande maioria das pesquisas, com base na literatura, entretanto, a aplicação destes itens não é feita por meio de testes com usuários, ou qualquer tipo de avaliação das preferências do usuário. Para isso, a elaboração do grupo de itens foca a objetividade.
- A abordagem seguida neste trabalho centra-se especificamente no site em si, não sendo consideradas as demais etapas de um processo de compra em websites tais como pós-venda, e etapas offline (logística, publicidade, etc).
- A coleta de dados é feita apenas em website de língua portuguesas, com o objetivo de evitar interferências linguísticas ou culturais.
- O foco de estudo são os websites comerciais. Portanto, questões ligadas exclusivamente à websites gerais baseados em publicidade não serão consideradas.
- O presente estudo trabalha particularmente sob o ponto de vista de sistema, ou seja, de projeto de websites ao invés de percepção de satisfação do usuário final, apesar de ter como base trabalhos anteriores que utilizam teste com usuários e percepção de usuários de sistemas web.

## 1.6 ESTRUTURA DO TRABALHO

Este trabalho foi estruturado levando-se em consideração as atividades necessárias para alcançar seus objetivos definidos na seção 1.3 configurando assim uma investigação planejada, desenvolvida e redigida de acordo com as normas acadêmicas.

O capítulo 2 apresenta uma revisão da literatura sobre conceitos relacionados à qualidade na web, criação de escalas de medidas e teoria da

resposta ao item mais especificamente votado para o estudo da dimensionalidade e adequação dos modelos.

No capítulo 3 é apresentado método de pesquisa utilizado, ou seja, o passo-a-passo utilizado para o desenvolvimento prático do trabalho. Nesta etapa é caracterizada a pesquisa, a elaboração dos itens a coleta de dados e é detalhada a metodologia estatística.

O capítulo 4 descreve o processo de avaliação dos resultados obtidos.

No capítulo 5 são apresentadas as considerações finais juntamente com as conclusões e as recomendações para trabalhos futuros.

Por último, são apresentadas as referências bibliográficas utilizadas para o desenvolvimento do trabalho.

## 2 REFERENCIAL TEÓRICO

Este capítulo aborda os conceitos básicos de qualidade em websites, escalas de medidas e teoria da resposta ao item (TRI). Dentro da conceituação de qualidade em websites são levantadas considerações sobre a definição de qualidade na web e é realizada uma revisão da literatura sobre as abordagens mais relevantes neste contexto. Na seção destinada aos conceitos de escalas de medida são abordadas suas tipologias, vantagens e desvantagens inseridas no contexto da web. Na sequência a conceituação da Teoria da Resposta ao Item é feita juntamente com um levantamento histórico desta ferramenta, alguns de seus modelos unidimensionais e multidimensionais, questões críticas para sua aplicação e nesta subseção são discutidas também algumas questões relacionadas à dimensionalidade.

### 2.1 QUALIDADE NA WEB

Reflexões a partir de estudos sobre a interação das pessoas com a tecnologia implicam que a avaliação dos clientes das novas tecnologias é um processo distinto do tradicional (PARASURAMAN et al., 2005) pois envolve um modo diferente de fazer negócios comumente chamado de *e-service*. Rust (2001) define *e-service* como o fornecimento de serviço utilizando redes eletrônicas. A globalização e o fácil acesso aos meios eletrônicos, conseqüentemente geram a este tipo de negócio um ambiente altamente competitivo. Sendo assim, para os fornecedores de *e-service*, a qualidade é uma das principais forças no caminho para o sucesso em longo prazo. Medidas de qualidade, confiáveis e detalhadas são a chave para a gestão eficaz destes serviços (FASSNACHT; KOESE, 2006).

A qualidade do serviço é um construto abstrato que é difícil de captar e medir (CRONION; TAYLOR, 1992). Entretanto, de uma forma geral, muitos pesquisadores consideram a qualidade do serviço uma medida de quão bem o nível de serviço prestado atende a expectativa do cliente (DING et al., 2011).

O instrumento mais aceito e amplamente adotado para medir a qualidade nos serviços tradicionais é o modelo SERVQUAL. Como serviço tradicional entende-se à qualidade de todas as interações com os clientes não baseadas na Internet e experiências diretas com as empresas (PARASURAMAN et al., 2005). Os primeiros escritos acadêmicos sobre qualidade em serviço (GRÖNROOS, 1984; PARASURAMAN et al., 1985)



sugerem que qualidade de serviço deriva de uma comparação entre o que os clientes sentem que uma empresa deve oferecer, isto é, suas expectativas, com o desempenho efetivo do serviço. Usando ideias desses estudos como ponto de partida, Parasuraman et al. (1988) realizaram estudos empíricos em diversos setores da indústria para desenvolver o modelo SERVQUAL, o qual mede a qualidade de serviço em cinco dimensões: tangibilidade, confiabilidade, receptividade, garantia e empatia.

- Tangibilidade: trata de elementos físicos e relaciona-se com avaliações dos clientes sobre as instalações, equipamentos e aparência dos prestadores do serviço.

- Confiabilidade: é a percepção do cliente de que o prestador do serviço está fornecendo o serviço prometido de forma confiável e segura, e está fazendo isso na hora.

- Receptividade: trata-se da dimensão relacionada as percepções dos clientes sobre a vontade do prestador de serviços em ajudá-lo e dar suporte aos seus pedidos de assistência.

- Garantia: conhecimento e cortesia dos funcionários e sua capacidade para inspirar confiança e segurança.

- Empatia: Cuidado e atenção individualizada prestadas aos clientes.

O SERVQUAL foi testado em vários contextos, tais como lojas físicas, saúde, turismo, indústria automotiva e sistemas de informação (LI et al., 2002).

Segundo Parasuraman et al. (2005), três conclusões gerais que são potencialmente relevantes para definir, conceituar e medir qualidade em serviços on-line (*e-service*), emergem da literatura tradicional de medida de qualidade em serviços:

- (a) a noção de que a qualidade do serviço decorre de uma comparação do desempenho do serviço real com o que deveria ou poderia ter sido conceitualmente entregue;

- (b) as cinco dimensões do SERVQUAL: tangibilidade, confiabilidade, receptividade, garantia e empatia captura o domínio geral de qualidade em serviços razoavelmente bem, apesar de ainda restarem dúvidas sobre se de fato as cinco dimensões são distintas e suficientes. Segundo Llosa et al. (1998), existem evidências empíricas de que confiabilidade, garantia e receptividade podem representar apenas um fator; e

- (c) as avaliações dos clientes sobre a qualidade de um serviço são fortemente ligadas ao valor percebido e as intenções comportamentais.

Entretanto, nota-se que as abordagens de qualidade em serviços tradicionais enfatizam as interações pessoais, ou seja, a avaliação das

pessoas envolvidas no processo, o que representa uma natureza diferente da qualidade em serviços online, na qual a interação entre pessoas não é o foco. Desta forma, verifica-se que avaliação de serviços online possui algumas peculiaridades em relação à avaliação de serviços tradicionais. Um dos estudos que evidenciou estas peculiaridades foi o de Gefen (2002), no qual o autor faz uma aplicação do SERVQUAL em empresas de comércio eletrônico (e-commerce) e conclui que as cinco dimensões originais do instrumento se reagrupam em três: (a) tangibilidade; (b) uma combinação das dimensões de receptividade, confiabilidade e garantia; e (c) empatia. Esta conclusão vai de encontro ao estudo de Llosa et al. (1998). O estudo de Gefen (2002) adaptou os itens para o contexto de e-commerce e concluiu que a dimensão de tangibilidade (aparência do website) é a mais relevante para aumentar a lealdade do cliente neste contexto.

Gefen (2002), Zeithaml et al. (2000), DeLone e McLean (2003) dentre outros têm se concentrado no desenvolvimento e adaptação de modelos para avaliação de serviço online. Entretanto, segundo Serkan et al. (2010), em comparação com a pesquisa abundante no contexto tradicional de qualidade em serviços, as pesquisas sobre a qualidade do serviço online ainda estão no começo.

Do ponto de vista técnico, as avaliações em serviços online receberam também influências dos estudos voltados para avaliação de sistemas de informação.

Na área de sistemas de informação, qualidade não é um conceito novo (ALADWANI; PALVIA, 2002). Do início da década de 1980 até o final da década de 1990 é possível encontrar na literatura vários estudos que tentam conceituar a qualidade em sistemas de informação, demonstrando a preocupação de profissionais e acadêmicos em compreender e melhorar estes sistemas. Estes estudos atinham-se em conceituar tópicos específicos neste contexto, tais como qualidade do gerenciamento de dados (KAPLAN et al, 1998, WANG et al, 1995), qualidade da informação (KING; EPSTEIN, 1983, HAGA; ZVIRAM, 1994), qualidade de softwares (ISO9126, 1992; SCHNEIDEWIND, 1992, KITCHENHAM; PFLEEGER, 1996), qualidade global do sistema (KETTINGER; LEE, 1994, NELSON, 1996), entre outros. Estes estudos eram muito mais focados na avaliação do sistema, seu desempenho e sua relação com usuários específicos.

A partir de meados da década de 1990, com o desenvolvimento e popularização da internet, práticos e pesquisadores têm se esforçado para definir qualidade no contexto da internet (por exemplo: BARNES; VIDGEN, 2000, DAY, 1997, LINDROOS, 1997, XIE et al., 1998, LOIACONO et al., 2002). Lindroos (1997) discute as diferenças entre os sistemas de informação baseados na *web* e sistemas de informação

convencionais a partir da perspectiva de qualidade de software. Olsina et al. (1999), com base na opinião de usuários, propõem um modelo de qualidade para sites de universidades, chamado Website QEM. Barnes e Vidgen (2000), Loiacono et al. (2002), Parasuraman et al. (2005) e Ding et al. (2011) também desenvolvem modelos semelhantes, mas focados em sites comerciais. Estes e vários outros estudos decompõem a qualidade dos websites em vários atributos. A criação de tais modelos é baseado principalmente em experiências de muitos anos no desenvolvimento e manutenção de sistemas de informação e web. A validação destes modelos é feita principalmente por estudos empíricos, tais como a análise dos dados coletados em testes com usuários, questionários de satisfação e entrevistas. Entretanto, diferentes tipos de sistemas de informação podem ter diferentes requisitos de qualidade (WORWA; STANIK, 2010). Por exemplo, websites comerciais e websites pessoais são sistemas de informação baseados na web, no entanto, os requisitos de qualidade são diferentes, principalmente em termos de segurança da informação e questões de busca de informações.

Segundo Worwa e Stanik (2010), o conceito de qualidade na web continua em desenvolvimento e ainda se apresenta de uma forma geral indefinida. Os autores ainda argumentam que a maior parte das pesquisas discute o significado de alguns aspectos de qualidade na web de forma descritiva, sem delinear as suas principais dimensões ou fornecer escalas interpretáveis. Além disso, nota-se que a maioria dos autores não define o domínio exato do construto qualidade, nem apresentam definição clara dos serviços online. Portanto, é importante definir explicitamente o que é serviço online ou e-service e o que significa qualidade neste contexto, a fim de dar uma noção mais clara do que e por que medir.

A falta de uma definição clara dificulta a direção da pesquisa e impede a comparação e a integração dos resultados. Segundo Fassnacht e Koese (2006), dimensões e subdimensões propostas em um artigo podem não ser sempre comparadas com outros estudos. Por exemplo, Gummerus et al. (2004) discutem a dimensão da qualidade da resposta em termos de feedback aos clientes, enquanto no trabalho de Loiacono et al. (2002), esta dimensão refere-se ao tempo de carregamento dos sites. Além disso, a grande maioria dos estudos é na realidade adaptação ou aplicação de modelos já existentes e, de acordo com Fassnacht e Koese (2006), a estrutura e o significado das dimensões elaboradas têm sido geralmente determinados *a posteriori* pelos resultados de análise de dados. Um referencial teórico consistente raramente tem sido usado. Alguns autores citam uma teoria específica ou conceito no início do desenvolvimento de seu modelo, o que não necessariamente se reflete na dimensionalidade da sua escala (GOUNARIS; DIMITRIADIS 2003; LOIACONO, et al., 2002).

Na tentativa de entender melhor os diversos estudos sobre qualidade em e-service, Cristobal et al. (2007) dividiram estes estudos em duas grandes categorias: serviços de varejo online e qualidade e design de website.

Entre os estudos na primeira categoria, destacam-se os modelos e-SERVQUAL desenvolvido por Zeithaml et al. (2000) e PIRQUAL desenvolvido por Francis e White (2002). No primeiro, Zeithaml et al. (2000) analisaram porque a maioria das empresas falham nas respostas às necessidades de seus clientes, segundo os autores, as razões para o fracasso estão relacionadas ao mau gerenciamento interno do negócio. Neste modelo, Zeithaml et al. (2000) identificaram 11 dimensões de qualidade em e-service:

- Acesso: ao site ou a empresa quando necessário;
- Garantia/confiança: o cliente se sente confiante ao lidar com o site;
- Facilidade de navegação: mudança rápida e facilmente por meio das páginas do site;
- Eficiência: site é simples de usar, requer o mínimo de introdução de dados por parte do cliente;
- Flexibilidade: opções para realizar uma transação eletrônica;
- Customização/personalização: com base nas preferências do cliente e histórico de compras;
- Conhecimento dos preços: sobre o preço individual, total e a possibilidade de comparar preços;
- Segurança/privacidade: site é seguro contra invasões, informações pessoais são protegidas;
- Estética do site: atributos de aparência;
- Confiabilidade: funcionamento técnico correto do site, cumprimento das promessas feitas ao cliente; e
- Receptividade: resposta rápida às necessidades do cliente.

Mais tarde, em Parasuraman et al. (2005), este modelo foi revisto e dividido em 2 escalas chamadas de E-S-QUAL e E-RecS-QUAL a primeira com 4 dimensões:

- Eficiência: acesso e utilização do local facilmente e rapidamente;
- Cumprimento: mantendo promessas sobre a entrega dos pedidos e disponibilidade de itens;
- Disponibilidade do sistema: funcionamento técnico correto do site; e
- Privacidade: site é seguro, as informações dos clientes são protegidas.

e a segunda, com 3 dimensões:

- Receptividade: tratamento eficaz de problemas;
- Compensação: site compensa clientes por problemas; e
- Contato: assistência via telefone ou representantes on-line.

O estudo de Francis e White (2002) também tenta abranger o negócio como um todo, gerando um instrumento com 23 itens em uma escala de 7 pontos (likert). Em seu estudo, os autores encontram 6 dimensões: (a) funcionalidade do website; (b) descrição dos atributos dos produtos; (c) condições de entrega; (d) entrega do produto; (e) serviço ao consumidor; e (f) segurança.

A segunda categoria ("qualidade e design de website"), sugerido por Cristobal et al. (2007), é baseada no fato de que o design do website é claramente a dimensão mais comumente examinadas nos estudos deste domínio (STIAKAKIS; GEORGIADIS, 2009). Exemplos desta categoria são os estudos de Loiacono et al. (2000), Barnes e Vidgen (2002), Wolfinbarger e Gilly (2003) entre outros que focam a qualidade do website. No caso de Loiacono et al. (2000), estes desenvolveram uma escala global, com 36 itens dentro de 12 dimensões: (a) informações ajustadas à tarefa, (b) comunicações personalizadas, (c) confiança, (d) tempo de resposta, (e) facilidade de compreensão, (f) intuitividade das operações, (g) recursos visuais, (h) inovação, (i) apelo emocional, (j) consistência da imagem, (k) completado online, (l) vantagem relativa.

No entanto, deve notar-se que objetivo principal destas escalas é gerar informações ao nível de gerenciamento do website, mais voltado para questões de projeto ao invés de abranger o serviço como um todo, ou seja, parte *online* e parte *offline* (PARASURAMAN et al., 2005).

O Quadro 1 apresenta uma revisão da literatura com abordagens relacionadas à qualidade em e-service. A revisão tem como referência as bases de dados Scopus, ScienceDirect, EBSCO e Emerald. Sendo que somente abordagens com foco no desenvolvimento de instrumentos para medir qualidade em e-service foram incluídos, para um período de 2000 à 2012.

Quadro 1- Classificação dos estudos mais relevantes sobre qualidade em e-service, de 2000 à 2012.

<b>Abordagem</b>	<b>Amostra</b>	<b>Método de análise</b>	<b>Nº de itens iniciais</b>	<b>Nº de itens finais</b>	<b>Dimensões</b>
Liu e Arnett, (2000)	119 profissionais	Análise fatorial	28	28	4 (qualidade da informação e do serviços, utilização do sistema, divertimento e qualidade do design do sistema)
O’Niell et al.(2001)	269 estudantes	Análise fatorial exploratória	18	18	4 (contato, confiabilidade, agilidade e tangíveis)
Yoo e Donthu (2001)	116 usuários	Análise fatorial exploratória e confirmatória	54	9	4 (facilidade de uso, projeto estético, velocidade de processamento e segurança)
Aldwani e Palvia (2002)	228 estudantes	Análise fatorial exploratória	55	25	4 (adequação técnica, conteúdo específico, qualidade do conteúdo e aparência do site)
Barnes e Vidgen (2002)	376 estudantes	Análise fatorial exploratória	22	22	5 (usabilidade, design, informação, confiança e empatia)
Francis e White (2002)	302 usuários	Análise fatorial exploratória	35	23	6 (funcionalidade do site, descrição dos atributos dos produtos, condições de entrega, entrega do produto, serviço ao consumidor e segurança)
Janda et al.(2002)	446 usuários	Análise fatorial confirmatória	30	22	5 (desempenho, acesso, segurança, sensação, e informação)
Li et al. (2002)	202 usuários e profissionais	Análise fatorial exploratória	28	25	6 (receptividade, competência, qualidade da informação, empatia, assistência online e sistema de pós-venda)
Loiacono et. al.(2002)	1154 estudantes	Análise fatorial exploratória e confirmatória	88	36	12 (informações ajustadas à tarefa, comunicações personalizadas, confiança, tempo de resposta, facilidade de compreensão, intuitividade das operações, recursos visuais, inovação, apelo emocional, consistência da imagem, completudo online e vantagem relativa)
Madu e Madu (2002)	Conceitual	Conceitual	-	-	15 (desempenho, características, estrutura, estética, segurança, capacidade de armazenagem, acessibilidade do serviço,

Abordagem	Amostra	Método de análise	Nº de itens iniciais	Nº de itens finais	Dimensões
					segurança e integridade do sistema, confiabilidade, responsividade, customização, políticas do site, reputação, auto-confiança e empatia)
Rangana than e Ganapathy (2002)	214 usuários	Análise fatorial exploratória	15	15	4 (conteúdo da informação, design, segurança e privacidade)
Yang e Jun (2002)	271 usuários	Análise fatorial exploratória	41	19	6 (confiabilidade, acesso, facilidade de uso, personalização, segurança e credibilidade)
Cai e Jun (2003)	171 estudantes	Análise fatorial exploratória	32	19	4 (conteúdo e design do site, confiabilidade, agilidade do serviço e comunicação)
Goenaris e Dimitriadis (2003)	603 usuários	Análise fatorial exploratória e confirmatória	14	13	3 (redução de riscos, benefícios da informação e facilidade de interação)
Santos (2003)	30 grupos focais	Qualitativo	-	-	5 (facilidade de uso, aparência, ligação, estrutura e layout e conteúdo)
Wolfinger e Gilly (2003)	1013 usuários	Análise fatorial exploratória e confirmatória	40	14	4 (design do website, confiabilidade, segurança/privacidade e serviço ao consumidor)
Gummerus et al. (2004)	421 usuários	Análise fatorial exploratória e confirmatória	10	6	4 (Interface com o usuário, responsividade, cumprimento dos prazos e segurança)
González e Palacios (2004)	200 sites	Análise fatorial exploratória	20	20	4 (acessibilidade, velocidade, navegabilidade e conteúdo)
Hung e McQueen (2004)	Artigo conceitual	Artigo conceitual	-	-	4 critérios (facilidade de identificação, facilidade de uso, utilidade da informação e interatividade)
Jun et al. (2004)	137 estudantes e profissionais	Análise fatorial exploratória	40	21	6 (confiabilidade/tempo de resposta, atencividade, facilidade de uso, acesso, segurança e credibilidade)
Kim e	273	Análise	36	25	6 ( aparência, entretenimento,

<b>Abordagem</b>	<b>Amostra</b>	<b>Método de análise</b>	<b>Nº de itens iniciais</b>	<b>Nº de itens finais</b>	<b>Dimensões</b>
Stoel (2004)	usuários	fatorial exploratória e confirmatória			informação à tarefa, capacidade de transação, tempo de resposta e confiança)
Long e McMellon (2004)	447 usuários	Análise fatorial exploratória	53	19	5 (tangibilidade, segurança, confiabilidade, processo de compra e responsividade)
Yang et al. (2004)	235 usuários	Análise fatorial confirmatória	27	20	6 (confiabilidade, responsividade, competência, facilidade de uso, segurança e portfólio de produtos)
Lee e Lim (2005)	297 estudantes	Análise fatorial confirmatória	15	15	5 (design do website, confiança, responsividade, confiabilidade e personalização)
Parasuraman et al. (2005)	1407 usuários	Análise fatorial exploratória e confirmatória	113	22	7 (eficiência, disponibilidade do sistema, cumprimento do prometido, privacidade, responsividade, compensação e contato)
Yang et al. (2005)	1992 usuários	Análise fatorial exploratória e confirmatória	37	19	6 (usabilidade, utilidade de conteúdo, adequação da informação, acessibilidade e interação)
Bauer et al. (2006)	384 usuários	Análise fatorial exploratória e confirmatória	53	25	5 (funcionalidade/ design, divertimento, processo, confiabilidade e responsividade)
Collier e Bienstock (2006)	600 estudantes	Análise fatorial exploratória e confirmatória	99	54	3 (processo, resultados e recuperação)
Fassnacht e Koese (2006)	999 usuários	Análise fatorial exploratória e confirmatória	36	24	Modelo hierárquico com 3 dimensões ( qualidade do ambiente, qualidade da entrega e qualidade dos resultados)
Ibrahim et al. (2006)	135 usuários	Análise fatorial exploratória	26	25	6 (conveniência, acessibilidade, gerenciamento dos serviço, personalização, amigabilidade e serviço alvo)
Cristobal et al. (2007)	461 usuários	Análise fatorial exploratória e confirmatória	86	17	4 (serviço ao consumidor, design, segurança e gerenciamento do pedido)
Ho e Lee	671	Análise	30	18	5 (qualidade da informação,



<b>Abordagem</b>	<b>Amostra</b>	<b>Método de análise</b>	<b>Nº de itens iniciais</b>	<b>Nº de itens finais</b>	<b>Dimensões</b>
(2007)	usuários	fatorial exploratória e confirmatória			segurança, funcionalidade do site, relacionamento com o consumidor e responsividade)
Nusair e Kandamully (2008)	6 websites	Análise de conteúdo	53	53	6 (navegabilidade, divertimento, qualidade de informação, confiança, personalização e responsividade)
Shachaf et al. (2008)	676 estudantes	Análise de variância	10	10	3 (tempo de resposta, confiabilidade e cortesia)
Sohn e Tadisina (2008)	204 usuários	Análise fatorial exploratória e confirmatória	30	25	6 (confiança, comunicação customizada, facilidade de uso, funcionalidade e conteúdo, confiabilidade e tempo de entrega)
Chang e Chen (2009)	314	Análise fatorial confirmatória e Modelagem de equações estruturais	27	27	8 (conveniência, interatividade, customização, características, percepção de segurança, custos, satisfação do consumidor e lealdade do consumidor)
Liu et al. (2009)	421 usuários	Análise fatorial exploratória e confirmatória	24	14	4 (usabilidade, privacidade e segurança, adequação da informação e aparência)
Ding et al. (2011)	311 estudantes	Análise fatorial exploratória e confirmatória	21	13	4 (percepção de controle, conveniência, serviço ao consumidor e prazo de entrega)
Hasan; Abuelrub (2011)	Artigo conceitual	Artigo conceitual	69	69	4 (conteúdo, design, organização e amigabilidade)
Lee; Kozar (2012)	689 usuários	Análise fatorial confirmatória e equações estruturais	43	38	10 (interatividade, telepresença, navegabilidade, aprendabilidade, legibilidade, suportabilidade, simplicidade, consistência, relevância do conteúdo e credibilidade)
Kim et al 2012	293 usuários e estudantes	Equações estruturais	48	45	6 (Segurança, Acessibilidade, Variedade, informação, velocidade, receptividade,

O Quadro 1 demonstra que os estudos em qualidade em serviços eletrônicos e websites vêm avançando com o tempo, estruturas teóricas tem sido adicionadas, confirmadas e/ou substituídas. No entanto, verifica-se que não existe um consenso teórico quanto à definição de qualidade em e-service e suas dimensões. Pode-se encontrar estudos com até 15 fatores e outros com 3 fatores, além disso alguns destes estudos contam com uma escala de 54 itens e outros com 9 itens, evidenciando a importância dada à interpretação do construto teórico frente aos resultados obtidos, bem como a metodologia utilizada.

Os estudos pesquisados utilizam variadas metodologias qualitativas (DING et al., 2011; CRISTOBAL et al., 2007; LIU et al., 2009; SANTOS, 2003; ZEITHAML et al., 2000), quantitativas (BAUER et al., 2006, GONZÁLEZ; PALACIOS, 2004) ou uma mistura de qualitativa com quantitativa (BARNES; VIDGEN, 2002; WOLFILNBARGER; GILLY, 2003; YANG et al., 2004). O tipo de metodologia mais frequente é a qualitativa, na qual autores como Ding et al. (2011), a partir de uma pesquisa sob percepção de 311 usuários, encontram 4 dimensões (i) percepção de controle, (ii) conveniência; (iii) serviço ao consumidor; (iv) cumprimento do serviço. Santos (2003) também utilizando método quantitativo conduziu 30 grupos focais para identificar 5 dimensões. Cristobal et al. (2007) utilizam a técnica qualitativa de entrevista em profundidade, juntamente com entrevista semi-estruturada. Bauer et al. (2006) e González e Palacios (2004) desenvolveram ferramentas automáticas para avaliar questões técnicas e quantitativas de websites. As metodologias mistas geralmente empregam análise de conteúdo e pesquisas com percepção de usuários como é o caso de Barnes e Vidgen (2002).

Ainda dentro da metodologia, pode-se perceber que a maioria das abordagens utilizou para análise dos dados a técnica de análise fatorial exploratória ou confirmatória, demonstrando que o desenvolvimento e/ou aplicação de novas técnicas pode representar um avanço técnico-metodológico neste campo.

Grande parte das amostras utiliza estudantes e o método de amostragem por conveniência (CAI; JUN, 2003; LEEAND; LIN, 2005; LOIACONO et. al. 2002; LONGAND; MCMELLON, 2004), poucas utilizam amostragem aleatória (DING et al., 2011; CRISTOBAL et al., 2007; FASSNACHT; KOESE, 2006; PARASURAMAN et al., 2005). Kim e Stoel (2004) e Ladhari (2010) criticam a utilização de estudantes nos estudos para criação de escalas, pois, segundo eles, esta pode ser uma deficiência do ponto de vista estatístico que pode comprometer sua generalização, uma vez que os estudantes, na maioria das vezes não representam realmente a população de usuários de e-commerce, sendo

muitas vezes convidados a participar da pesquisa como complemento de seus estudos. Dentro das amostragens aleatórias, existem também algumas limitações, por exemplo, Parasuraman et al. (2005) utilizaram somente respondentes que utilizaram internet nos últimos 12 meses e que tivesse feito pelo menos 3 compras durante os últimos 3 meses. Segundo Yang et al. (2004) limitações como as impostas por Parasuraman et al. (2005) são importantes tendo em vista que usuários que têm a internet exclusivamente como fonte de informação possuem uma percepção diferente dos usuários que utilizam a internet para realizar transações comerciais, principalmente com relação à segurança. Alguns autores fazem a coleta de dados via e-mail; por exemplo, Kim e Stoel (2004) e Sohn e Tadisina (2008). Entretanto, uma desvantagem deste método é que a taxa de respostas geralmente é baixa. Kim e Stoel enviaram 1000 formulários e obtiveram resposta de apenas 273. Sohn e Tadisina (2008) misturam a coleta via e-mail com aplicação direta em 2000 usuários selecionados e obtiveram 204 respostas.

Outra questão relevante do ponto de vista metodológico dos estudos levantados é a grande diferença na quantidade da amostra. Por exemplo, enquanto Jun et al. (2004) utilizam 137 usuários (estudantes e profissionais), Yang et al. (2005) utilizam 1992 usuários. Como a grande maioria das abordagens pesquisadas utiliza como procedimento para análise dos dados a análise fatorial, pode-se comparar a confiabilidade das escalas propostas. No caso da escala proposta por Jun et al. (2004), a confiabilidade (alpha de Cronbach) ficou entre 0,59 e 0,92, enquanto que no estudo de Yang et al. (2005), a confiabilidade variou de 0,66 a 0,89. Evidentemente que o valor da confiabilidade interna de uma escala não depende exclusivamente do tamanho da amostra, mas, também da consistência dos itens e o número de itens. Geralmente, assume-se como aceitável valores acima de 0,70 um valor aceitável, mas, segundo Kline (2000), em casos com grande diversidade dos construtos como, por exemplo, testes psicológicos, é aceitável valores abaixo de 0,70. A maioria dos estudos pesquisados apresentou um alpha de Cronbach maior que 0,70 com exceção do estudo de Ibrahim et al. (2006), o qual apresenta uma amplitude de 0,33 a 0,84.

Apesar das deficiências evidenciadas nestes estudos, é inegável que eles representam um considerável avanço na exploração de medidas de qualidade em websites, indicando que existe um vasto campo a ser avançado e várias lacunas a serem preenchidas com mais estudos tanto do ponto de vista teórico como metodológico.

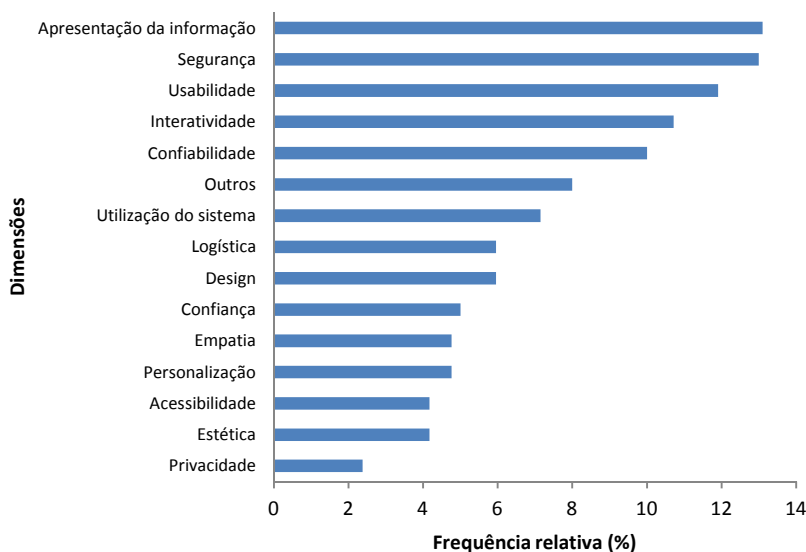
O presente estudo se encaixa na classificação de Cristobal et al. (2007) como um estudo da qualidade e design de website. Dentro deste

escopo, assume-se qualidade em websites como a qualidade de um sistema de informação que, segundo Loiacono et al. (2002), foca no armazenamento, processamento, apresentação e transferência da informação.

Sendo assim, o conceito adotado de qualidade em websites é o de um conjunto de características técnicas e não técnicas de um sistema web, os quais permitem ao usuário realizar seus objetivos em um website de forma acessível, eficiente e agradável. Como características técnicas entende-se usabilidade/navegabilidade, apresentação da informação, acessibilidade e interatividade do sistema. Como características não-técnicas entende-se design, estética, apelo visual/comercial, confiabilidade, hedonismo, empatia.

Tal definição tem como base as abordagens dos estudos apresentado no Quadro 1, particularmente os resultados referentes as definições das dimensões. A Figura 1 apresenta um resumo das dimensões discutidas nos artigos levantados no Quadro 1, tomando como base a frequência com que estas dimensões são abordadas.

Figura 1- Distribuição de frequência das abordagens levantadas no Quadro 1 quanto às dimensões mais citadas



Verifica-se na Figura 1 que a dimensão relacionada com a apresentação da informação juntamente com segurança e usabilidade são as dimensões mais encontrada nas abordagens levantadas. Dentre as dimensões mais citas, com exceção da dimensão logística, que se refere a prazos de entrega, meios de entrega e/ou pós-venda, todas possuem relação direta com o sistema web, portanto ligadas diretamente a definição de qualidade em websites adotada anteriormente. A categoria outros está ligada a conceitos/dimensões particularmente pontuais, como por exemplo, divertimento (LIU; ARNETT, 2000; BAUER et al., 2006 e NUSAIR; KANDAMPULLY, 2008), portfólio de produtos (YANG et al., 2004) agilidade do serviço (O'NEILL et al., 2001 e CAI; JUN, 2003), credibilidade (LEE; KOZAR, 2012 e JUN et al., 2004).

Dado este levantamento bibliográfico sobre as técnicas de avaliação de qualidade mais comuns na literatura e os resultados encontrados juntamente com o direcionamento conceitual do termo qualidade em websites comerciais, aborda-se nas próximas seções o conceito de criação de escalas e os conceitos básicos de avaliação de dimensionalidade e de alguns modelos da teoria da resposta ao item.

## 2.2 CRIAÇÃO DE ESCALAS

Escalas são instrumentos que objetivam medir traços latentes, intensidade de opiniões ou atitudes da maneira mais objetiva possível. Uma escala consiste em um arranjo, em forma de série graduada de itens, pelo qual se mede uma característica, de acordo com um número previamente determinado e são compostas por categorias, que são as gradações ou alternativas de respostas oferecidas. O que se espera de uma escala é a discriminação de medidas. (GIL, 2002; MOTTA, 1999).

A construção de escalas de medida facilita o entendimento de conceitos complexos. Três razões explicam isso. Primeiramente, procurar por vários aspectos de uma variável fornece a criação de conhecimento sobre ela. Segundo, várias perspectivas realçam o intervalo de variação, o que permite distinções mais precisas, principalmente se envolver medidas ordinais. Em terceiro lugar, a construção de medidas permite uma redução eficiente dos dados por meio da qual uma pontuação numérica pode representar a posição ordinal onde determinada característica do item ou de elementos da população em avaliação se encontram na escala, o que permite comparabilidade (BABBIE, 2005).

Para construir uma escala baseada num conjunto de itens, as variáveis devem estar ligadas teoricamente ou conceitualmente com o que a pesquisa visa medir (BAKER, 2009).

A maior implicação das escalas é a possibilidade de comparação entre escores e apropriadas estatísticas que resumem estes escores. Em uma escala de medida, o papel dos números é de representar um objeto, ou uma pessoa, de forma que a relação entre os números represente a relação empírica entre os objetos ou entre as pessoas. Esta representatividade fornece ao objeto importantes propriedades tais como, poder de distinção entre objetos, ordem, adição e razão (EMBRETSON; REISE, 2000).

### 2.2.1 Tipologias de escalas

A distinção entre tipos de escala pode ser realizada por meio de quatro níveis de medidas: escala nominal, escala ordinal, escala intervalar, escala de razão (EMBRETSON; REISE, 2000; GÜNTHER, 2003; PASQUALI, 1997; SOMMER; SOMMER, 1997).

Na escala nominal, números ou símbolos são utilizados para representar categoria de objetos ou pessoas. Números específicos e símbolos são simplesmente rótulos, propriedades que definem ordem adição ou razão não são atingidas em medições nominais (EMBRETSON; REISE, 2000).

Numa escala ordinal, além de identificarem pessoas, objetos ou categorias, números ou símbolos os ordenam numa dimensão subjacente. Um exemplo da aplicação desta escala é a hierarquização de preferências ou importâncias entre pessoas ou objetos, status social propriedades de um sistema etc (GÜNTHER, 2003). Dados de uma escala ordinal satisfazem todas as propriedades de dados nominais e adicionam a representação numérica uma ordem as pessoas ou ao traço latente. Permite transformações e preservam tanto a distinção dos dados quanto a ordem (EMBRETSON; REISE, 2000).

Na escala intervalar, as características não somente podem ser ordenadas conforme uma dimensão subjacente, mas os intervalos entre as alternativas têm tamanho conhecido e podem ser comparados (GÜNTHER, 2003). Nesta escala são permitidas apenas transformações lineares, ou seja, escores podem ser multiplicados por uma constante e ou ter uma constante adicionada (por exemplo,  $aX + b$ ). Parâmetros estatísticos são apropriados para escalas intervalares, tais como média, desvio padrão e correlações (EMBRETSON; REISE, 2000).

Em uma escala de razão, a razão dos escores é significativa (por exemplo, um escore de 6 representa o dobro do escore 3 em determinada propriedade). A única transformação permitida neste tipo de escala é a multiplicação por uma constante que preserve a razão entre os escores (EMBRETSON; REISE, 2000).

A diferenciação entre os quatro tipos de escala tem consequências importantes quanto a complexidade da análise estatística desejável. Dados obtidos em qualquer escala podem ser apresentados por meio de estatísticas descritivas. Para utilizar estatísticas inferenciais, que permitem ao pesquisador verificar até que ponto determinadas relações ou diferenças são sistemáticas ou não, há de se observar que dados baseados em escalas nominais e ordinais podem ser trabalhados com testes estatísticos paramétricos e não paramétricos. Dados provenientes de escalas intervalares e de razão vão permitir, além de estatísticas não paramétricas, procedimentos paramétricos (GÜNTHER, 2003).

A escala Likert é mais utilizada nas ciências sociais e humanas, especialmente em levantamento de atitudes, opiniões e avaliações. Nela pede-se ao respondente que avalie um fenômeno numa escala de, geralmente, cinco pontos: aplica-se totalmente, aplica-se, nem sim nem não, não se aplica, definitivamente não se aplica. As afirmações podem ser autorreferentes: “Eu considero importante em um site ter acesso em outras línguas além do português”. Ou hétero-referentes: “É importante para a comunidade estrangeira ter acesso em outras línguas além do português”. Dependendo do tema subjacente, as alternativas podem, além da dimensão “aplica-se”, seguir dimensões como: “bom – ruim” ou concordo – discordo” (GÜNTHER, 2003). Gil (2002) acrescenta que a escala de Likert é de elaboração simples e tem caráter ordinal ou intervalar, e não mede quanto uma atitude é mais ou menos favorável. É uma escala onde os participantes registram sua concordância ou discordância com um enunciado.

### 2.2.2 Vantagens e desvantagens das escalas

A maior vantagem de uso de escalas é que são instrumentos estruturados e padronizados. Isso permite que a avaliação seja facilmente comparada e contrastada, mesmo para grande número de respostas. As escalas geralmente são fáceis de usar e de entender, o que justifica a ampla aceitação dessa abordagem. Em contraponto, as escalas também podem apresentar algumas desvantagens, como a possibilidade de ocorrerem erros de percepção e de significado percebido. Da mesma forma, a terminologia utilizada na construção da escala pode também influenciar a percepção do

respondente, pois termos como “excede as expectativas”, ou “capacidade abaixo da média” podem significar coisas diferentes para pessoas diferentes (SANTOS, 1999).

Outra desvantagem é a possibilidade de ocorrer um erro de avaliação, sendo a tendência central o mais comum. Respondentes ocupados podem tender a responder rapidamente e marcar as opções centrais de uma escala, como “satisfatório”, “adequado”, demonstrando uma atitude excessivamente passiva frente às questões colocadas. Ao se elaborar uma escala, alguns problemas básicos se apresentam: definição de um contínuo; confiabilidade; validade; ponderação e natureza de itens. (GIL, 2002; O’CONNOR, 2007)

A confiabilidade de uma escala se reflete nos resultados semelhantes e consistentes obtidos após a aplicação repetida com a mesma amostra. Para se verificar a confiabilidade de uma escala, podem ser aplicadas técnicas como a de teste-e-reteste, onde a escala é aplicada duas vezes à mesma população e se comparam os resultados. Neste método, a correlação entre as provas é um índice de confiabilidade ou de consistência. A validade de uma escala é observada quando ela mede exatamente aquilo a que se propõe medir. A fim de se verificar o grau de validade de uma escala, os itens podem ser submetidos ao julgamento de um grupo de pessoas consideradas especialistas em determinado domínio do conhecimento. Outra forma de verificar a validade de uma escala é submetê-la ao crivo de grupos conhecidos distintos, como votantes de esquerda e de direita. Se forem observadas distinções entre os resultados dos dois grupos, então a escala tem validade (SANTOS, 1999).

Outra fonte de problema em potencial está relacionada à ponderação dos itens. As formas utilizadas para se verificar a validade também podem ser aplicadas para se verificar a ponderação dos itens. Gil (2002) destaca, também, o problema relacionado à natureza dos itens, que podem ser diretos ou projetivos. Os itens diretos apelam para respostas diretas de concordância ou discordância, ao passo que os projetivos são apresentados como estímulos para as respostas, como frases incompletas, por exemplo.

### 2.2.3 Elaboração de um conjunto de itens

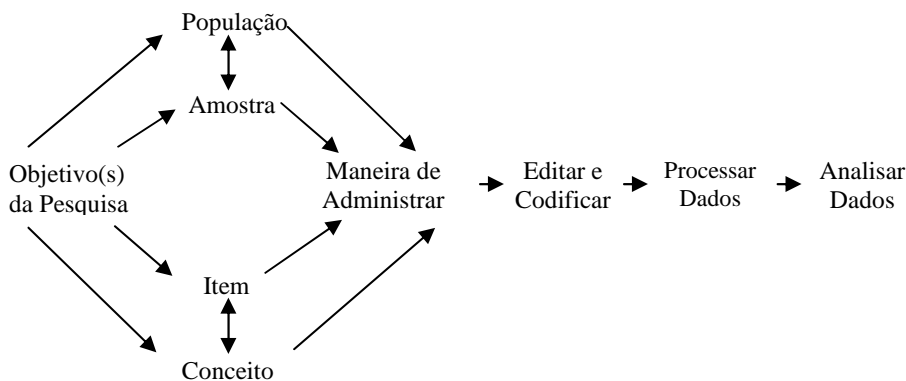
Conjunto de itens, ou questionários, é composto por uma quantidade razoavelmente elevada de questões apresentadas às pessoas, e uma forma de registrar suas respostas (GIL, 2002; MOTTA 1999). O objetivo é conhecer as opiniões e atitudes, de maneira objetiva, em relação



à interface em questão. Os questionários podem ser compostos, de maneira geral, por perguntas fechadas ou perguntas abertas. A existência de cada tipo de pergunta dependerá do tipo de dado a ser coletado. Em questões fechadas, o usuário especifica se concorda, se discorda ou se está indeciso, também especifica sua preferência ou concordância, e, ainda, pode escolher um item em uma lista. As perguntas abertas podem resultar em boas ideias, mas são mais difíceis de analisar e de quantificar. Sempre é necessário realizar um teste-piloto com o questionário, a fim de validá-lo. Um questionário para coletar dados referentes à satisfação do usuário precisa ser elaborado de maneira a possuir bom nível de confiabilidade (poder ser aplicado em situações similares e obter os mesmos resultados) e bom nível de validade (saber se são obtidos exatamente os dados necessários) (SANTOS, 1999).

A elaboração de um questionário, segundo Schuman e Kalton (1985) parte da definição do(s) objetivo(s) da pesquisa, que no presente trabalho é medir qualidade em sites de e-commerce. Este(s) objetivo(s) segundo os autores levam necessariamente em consideração a relação conceito/item e a relação população-alvo/amostra, como ilustrado na Figura 2.

Figura 2 - Principais estágios de um survey



Fonte: Adaptado de Schuman e Kalton (1985)

Conceito e população representam os termos abstratos da pesquisa, traduzidos em termos práticos pelo item e pela amostra.

O modelo proposto por Schuman e Kalton (1985) indica também uma relação direta entre população-alvo e amostra assim como entre

conceito e item, atrelados ao objetivo da pesquisa. Desta forma, dependendo do objetivo da pesquisa e da população-alvo, definir-se-ão diferentes tipos de amostra, o mesmo ocorre entre conceito e item.

Na sequência os conceitos, o conteúdo dos itens, a população-alvo e a amostra determinam o instrumento e a maneira de administrar. Segundo Günther (2004) a interdependência entre a elaboração de um instrumento e a estratégia de sua aplicação está nos seguintes pontos:

(1) o grau de complexidade dos conceitos determina o número de itens e a forma de apresentação deles;

(2) existe relação recíproca entre as características da população-alvo e a complexidade dos conceitos a serem investigados. Ambos determinam a maneira de transformação dos conceitos em itens e sua aplicação;

(3) o tamanho da amostra influencia a maneira de aplicar o instrumento, em termos de entrevista versus questionário e em termos de tamanho.

A utilização de item para avaliar qualidade em websites é uma técnica de análise comumente utilizada na literatura e no meio empresarial, principalmente para captar percepções de usuários.

Baker (2001) afirma que cada item de um conjunto de itens deve medir alguma faceta da habilidade particular de interesse, e do ponto de vista puramente técnico, cada item deve ser de resposta livre, para que o examinado possa escrever qualquer resposta que lhe possa ser apropriada.

## 2.3 TEORIA DA RESPOSTA AO ITEM

Teoria de Resposta ao Item (TRI) é composta de um conjunto de modelos probabilísticos, que relaciona um traço latente de um respondente ( $\Theta$ ), algo que não pode ser medido diretamente, com a probabilidade de este responder a um item dentro de uma determinada categoria (LORD, 1980). Os itens são geralmente classificados como dicotômicos o que significa que existem apenas duas categorias de resposta (normalmente denominadas "correto" e "incorreto"), ou politômicos, o que significa que as respostas podem ser classificadas em várias categorias (BROSSMAN, 2010).

O clássico livro de Lord e Novick (1968) é frequentemente citado na literatura como a base fundamental para o desenvolvimento moderno da TRI, principalmente na área educacional e psicológica. Entretanto, uma investigação histórica mais rigorosa demonstra que as bases conceituais são um pouco mais antigas, e surgem do questionamento de pesquisadores

como Thurstone (1927, 1928), no qual são colocados em dúvida os métodos de avaliação vigentes na época – teoria clássica dos testes (TCT). Na TCT as conclusões são fundamentadas sobre o escore total do teste, composto pelo conjunto de todos os itens, desta forma, esta teoria tem como objetivo a interpretação da resposta final, ou seja, o que a soma dos itens diz sobre o respondente, tornando o resultado pontencialmente dependente do conteúdo da prova (para mais detalhes sobre a TCT ver DeVellis, (2006)). Sendo assim, para Thurstone (1928), não é cientificamente admissível que os instrumentos de medida sejam seriamente afetados pelo objeto de medida.

Com o propósito de equacionar este problema, alguns autores tais como Richardson (1936), Mosier (1941, 1942), Ferguson (1942), Lawley (1943) e Lord (1952), forneceram alguns métodos para estimar os parâmetros dos itens, afastando-se teoricamente da teoria clássica dos testes. Uma revisão histórica mais detalhada da TRI pode ser encontrada em Baker e Kim (2004), Bock (1997), Embretson e Reise (2000), e van der Linden e Hambleton (1997).

Antes de apresentar alguns dos modelos matemáticos da TRI, é importante discutir alguns pressupostos básicos para utilização desta ferramenta.

Segundo Hambleton et al. (1991) há dois pressupostos fundamentais envolvidos na TRI. A primeira premissa afirma que uma mudança na variável latente, conduz a uma mudança no mesmo sentido na probabilidade de uma resposta específica que é descrita pela curva característica do item (CCI). Esta CCI especifica como a probabilidade de uma resposta ao item é alterada em função das mudanças na variável latente. Diferentes formas matemáticas da CCI conduzem a diferentes modelos de resposta ao item. Esta premissa é conhecida como monotonicidade.

A segunda premissa afirma que a resposta de qualquer pessoa para qualquer item do teste depende somente da habilidade ou conhecimento do indivíduo e dos parâmetros do item. A implicação prática desta premissa é que o examinado não divide informações durante o teste e as informações de um item não ajudam ou impedem o processo de resposta de outro item (RECKASE, 2009). Esta premissa é chamada de independência local ou condicional e é garantida quando o modelo correto é utilizado, ou seja unidimensional ou multidimensional. Desta forma, a suposição de independência local é facilmente generalizada para uma variável latente, multidimensional, neste caso, as respostas a um conjunto de itens são estatisticamente independentes quando a variável latente multidimensional é mantida constante (FOX, 2010).

Segundo Andrade et al. (2000) as vantagens da utilização da TRI dependem fundamentalmente da adequação (ajuste) dos modelos e seus pressupostos. Por exemplo, somente a partir de modelos com bom ajuste é que se pode garantir a obtenção de itens e habilidades invariantes, ou seja, não dependentes entre si. Reckase (2009) argumenta que estes pressupostos têm como função principal simplificar a manipulação matemática para aplicação dos modelos da TRI, e são particulares para modelos acumulativos, dicotômicos e com dimensionalidade bem definida (questões relacionadas a dimensionalidade serão discutidas mais a frente). A premissa de monotonicidade impõe limites às formas matemáticas consideradas para a função de resposta ao item, e a independência local permite uma grande simplificação na estimação dos parâmetros. Entretanto, é possível encontrar na literatura modelos não-monotônicos: Thissen e Steinberg (1984), Roberts et al. (2000), Roberts e Huang (2003) entre outros.

Desta forma, fica claro que a TRI apresenta algumas vantagens frente a outros instrumentos como, por exemplo, a TCT, entretanto, como qualquer método estatístico é necessário ter um conhecimento prévio da natureza do traço latente a que se pretende medir, para escolha do modelo mais adequado. Neste sentido alguns autores (EMBRETSON; REISE 2000, SCHUMACKER, 2005) enfatizam algumas vantagens e desvantagens na utilização da TRI, com base nestas premissas.

Como vantagens, os referidos autores destacam:

- As estatísticas dos itens são independentes da amostra a partir do qual foram estimadas;
- A pontuação dos respondentes é independente da dificuldade do teste;
- Análise do teste não exige rigorosos testes paralelos para avaliar a confiabilidade
- Item e respondentes são posicionados na mesma escala.

Schumacker (2005) descreve algumas limitação técnicas e práticas da TRI. Para este, os pressupostos para a utilização dos modelos TRI são mais rigorosas que as previstas à TCT. Modelos de TRI tendem a ser mais complexo e as saídas do modelo, mais difíceis de entender, especialmente para o público não técnico. Além disso, os modelos da TRI exigem grandes amostras para obter estimativas precisas e estáveis dos parâmetros. Consequentemente, a escolha de um modelo pode depender da amostra disponível, especialmente na fase de testes.

Como citado anteriormente, a TRI é composta de vários modelos matemáticos que se ajustam melhor dependendo dos interesses da pesquisa,

desta forma, no presente capítulo não serão tratados todos estes modelos, mas sim o grupo de modelos mais comuns na literatura, e algumas generalizações para dados multidimensionais.

### 2.3.1 Modelos unidimensionais da TRI

Segundo Reckase (2009) a representação básica do modelo unidimensional da TRI pode ser descrita pela seguinte equação:

$$P(U = u|\theta) = f(\theta, \eta, u) \quad (1)$$

Onde,  $\theta$  representa o parâmetro que descreve as características do respondente,  $\eta$  representa o vetor dos parâmetros que descrevem as características do item,  $U$  representa a resposta ou a pontuação em determinado teste, e  $u$  é uma possibilidade de resposta, em modelos dicotômicos 1 (resposta positiva) 0 (resposta negativa), e  $f$  é a função que descreve a relação entre os parâmetros e a probabilidade de uma resposta  $P(U = u)$ .

Um dos modelos mais simples da TRI é o modelo de um parâmetro ou modelo de Rasch (Rasch 1960). Neste modelo existe um parâmetro para o indivíduo  $\theta_j$  e um parâmetro para o item  $b_i$  que representa o grau de dificuldade do item, neste caso, a equação (1) ficaria desta forma:

$$P(U_{ij} = u_{ij}|\theta_j) = f(\theta_j, b_i, u_{ij}) \quad (2)$$

Onde  $u_{ij}$  é a resposta do respondente  $j$  ao item  $i$  (0 ou 1).

A parte esquerda desta equação representa a probabilidade da função  $f$  sendo assim, esta função necessita ter resultados entre 0 e 1. Além disso, para avaliações cognitivas, que eram o foco de interesse no desenvolvimento deste modelo, espera-se que o resultado seja cumulativo, em outras palavras, quanto maior a habilidade do respondente, maior a probabilidade deste responder a categoria esperada. Esta necessidade vai ao encontro ao pressuposto de monotonicidade.

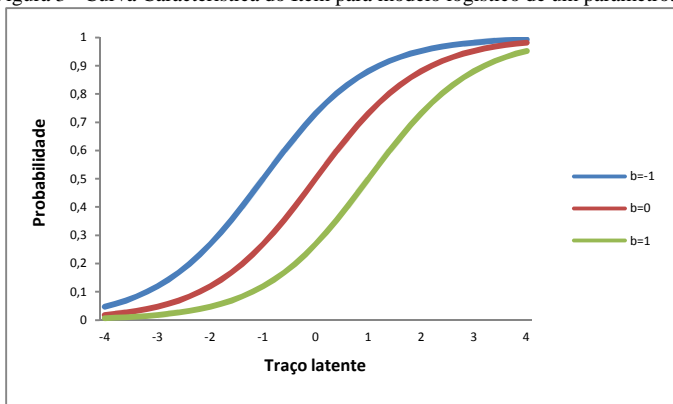
A representação da equação (2) como modelo logístico de um parâmetro ou modelo de Rasch é conhecida da seguinte maneira:

$$P(U_{ij} = 1|\theta_j) = \frac{e^{(\theta_j - b_i)}}{1 + e^{(\theta_j - b_i)}} \quad (3)$$

$b_i$  é o parâmetro de dificuldade do item, e representa a posição do item  $i$  na escala do traço latente, e  $\theta_j$  representa a posição do respondente  $j$  na mesma escala.

A Figura 3 mostra a CCI para três itens hipotéticos segundo este modelo.

Figura 3 - Curva Característica do Item para modelo logístico de um parâmetro.



Segundo Embretson e Reise (2000) existem algumas características gerais da CCI no modelo de um parâmetro. A primeira delas, como já foi discutido anteriormente na premissa da monotonicidade, a probabilidade aumenta gradualmente conforme aumenta o traço latente para cada item. Segundo, os itens se diferem somente na dificuldade, a inclinação da curva é igual para todos os itens e as curvas convergem, mas não se cruzam. Terceiro, o ponto de inflexão da CCI representa o ponto onde a probabilidade é 0,5.

A segunda característica, a qual afirma que a inclinação da curva é igual para todos os itens levantou algumas discussões quanto ao valor desta inclinação. Segundo a equação (3), este valor é igual a 1, mas teoricamente não há restrição para que este valor seja diferente de 1, para mais detalhes desta particularidade ver de Ayala (2009).

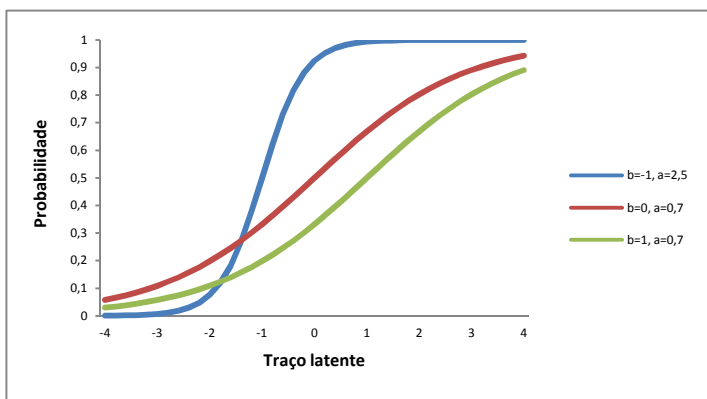
Em suma, o modelo de um parâmetro está dentro da premissa de que a distância entre a localização do respondente e a localização do item é um importante determinante de sua resposta. Entretanto, quando se examina o índice de discriminação tradicional, como, por exemplo, a correlação bisserial, encontra-se variação desta discriminação entre os itens (de

AYALA, 2009). A necessidade de considerar esta discriminação entre itens deu origem ao modelo logístico de dois parâmetros, representado da seguinte maneira:

$$P(X_{ij} = 1 | \theta_j) = \frac{1}{1 + e^{-a_i(\theta_j - b_i)}} \quad (4)$$

Em comparação com a equação (3), a equação (4) adiciona o parâmetro de discriminação do item, representado pelo coeficiente  $a_i$ . O modelo de dois parâmetros é apropriado para medições em que os itens não são igualmente relacionados como traço latente (EMBRETSON; REISE, 2000). A Figura 4 mostra a CCI para três itens hipotéticos segundo o modelo de dois parâmetros.

Figura 4 - Curva Característica do Item para modelo logístico de dois parâmetros.



Na Figura 4, é possível verificar que os itens têm a mesma dificuldade dos itens da figura 1, entretanto discriminações diferentes. Nota-se que a CCI do item com  $b=-1$  cruza os demais itens e dependendo a região do traço latente, este item apresenta uma maior distinção entre traço latente e probabilidade de uma resposta positiva.

O modelo mais geral para análise de itens unidimensionais com resposta dicotômica é o modelo logístico de 3 parâmetros. Neste modelo, um terceiro parâmetro é inserido, com o objetivo de modelar a resposta ao acaso, bastante comum em provas de conhecimento. A formulação matemática deste modelo é dada da seguinte forma:

$$P(U_{ij} = 1 / \theta_j) = c_i + (1 - c_i) \frac{1}{1 + e^{-a_i(\theta_j - b_i)}} \quad (5)$$

onde  $P(U_{ij}=1|\theta_j)$  representa a probabilidade de uma resposta positiva, e  $\theta_j$  representa o nível do traço latente,  $b_i$  representa a dificuldade do item  $i$  em uma determinada escala e o valor da variável latente  $\theta$ , para o qual há  $(1+c_i)/2$  de probabilidade do indivíduo  $j$  escolher a resposta representada por  $U = 1$ . O coeficiente  $a_i$  é o parâmetro de discriminação do item  $i$ , proporcional à inclinação da Curva Característica do Item – CCI no ponto  $b_i$ . O parâmetro  $c$  representa a probabilidade de acerto casual. Segundo Barker (2001), por definição, o valor de  $c$  não varia em função do nível da habilidade; assim, os respondentes com a mais baixa e a mais alta capacidade têm a mesma probabilidade de acerto casual.

Ao impor algumas restrições sobre os parâmetros, o modelo de 3 parâmetros pode ser transformado em dois modelos de menor complexidade (LORD, 1980). Quando não houver a possibilidade de resposta ao acaso, o parâmetro  $c$  é definido como zero, sendo simplificado para um modelo de dois parâmetros (ML2). Quando se assume  $c = 0$  e  $a = 1$ , simultaneamente, simplifica-se para o modelo de um parâmetro (ML1).

Entretanto, se as respostas aos itens forem politômicas, uma série de outros modelos é indicada. Para respostas politômicas ordinais, ou seja, quando a ordem das alternativas é importante, por exemplo, escala Likert pode-se utilizar os Modelos de Crédito Parcial - MCP (MASTERS, 1982; MASTERS; WRIGNT, 1997); Modelo de Crédito Parcial Generalizado – MCPG (MURAKI, 1992), Modelo de Resposta Gradual - MRG (SAMEJIMA, 1969), entre outros. Segundo Andrade et al., (2000) o MCP é uma extensão do modelo de Rasch ou modelo de um parâmetro para respostas politômicas. O MCPG é uma extensão do MCP no qual é adicionado o parâmetro de discriminação. O MRG assim como o MCPG assume que o parâmetro de discriminação do item, é o mesmo para todas as categorias do item. Bock (1972) propôs um modelo para categorias nomais, nas quais não importa a ordem da resposta. Estes modelos politômicos assumem que as categorias de resposta de um item podem ser ordenadas entre si e com isso, tenta obter mais informação das respostas dos indivíduos do que, simplesmente, se forneceram respostas corretas ou incorretas (ANDRADE et al., 2000). Para mais detalhes destes modelos ver van der Linden e Hambleton (1997), Embretson e Reise (2000), Andrade et al. (2000), de Ayala (2009). Para detalhes de modelos da TRI para traço latentes não cumulativos ver Andrich (1978), Roberts et al. (2000), Bortolotti (2010).



### 2.3.2 Dimensionalidade e Teoria da Resposta ao Item

Uma das suposições empregadas nos modelos apresentados na sessão anterior é a suposição da unidimensionalidade, a qual define que todos os itens do teste estão medido o mesmo traço latente ou a mesma composição de múltiplas habilidades. Entretanto, existem muitas situações em que os itens que compõem o teste individualmente, podem estar medindo diferentes dimensões do traço latente, ou diferentes composições de múltiplas habilidades (ACKERMAN, 1994). Um exemplo bastante comum na literatura é uma prova que envolve problemas de matemática. Neste caso pode-se ter, por exemplo, duas dimensões que determinam a proficiência do respondente. Uma dimensão reflete a proficiência em matemática e a outra reflete a proficiência em leitura e entendimento do problema.

Traub (1983) argumenta que quando se considera todas as habilidades necessárias para resolver todos os itens na maioria dos testes cognitivos, a unidimensionalidade é, provavelmente, mais a exceção do que a regra.

Modelos tradicionais da TRI, como os apresentados na seção 2.3.1 assumem que os dados podem ser explicados por um único traço latente. Entretanto, esta restrição torna estes modelos inapropriados para dados com estrutura multidimensional (HAMBLETON; SWAMINATHAN, 1985).

Do ponto de vista teórico, Reckase (1997) afirma que dependendo dos objetivos do estudo, a Teoria da Resposta ao Item Multidimensional (TRIM) pode ser considerada ou como um caso especial da técnica de Análise Fatorial (AF) ou Modelagem de Equações Estruturais (MEE), ou ainda como uma extensão de TRI unidimensional. Estas diferentes perspectivas dentro da TRIM afetam como a metodologia é aplicada e a interpretação de seus resultados. Esta seção tem como objetivo apresentar estes casos e discutir a metodologia empregada nestes e a interpretação dos resultados. Mas antes de discutir os modelos propriamente ditos, faz-se necessário uma breve discussão sobre dimensionalidade.

Qualquer objeto pode ser imaginado como tendo dimensões que representam as percepções de um indivíduo quanto a atributos ou combinações destes. Estes atributos podem ser classificados como quantificáveis (físicos ou observáveis), também chamados de dimensão objetiva ou perceptiva, também conhecida como dimensão subjetiva. Neste caso, a dimensão percebida é uma interpretação feita pelo indivíduo que pode ou não ser baseada em dimensões objetivas (HAIR et al., 2009).

De uma forma geral, dados que possam ser representados ou explicados por apenas uma dimensão são classificados como unidimensionais, por duas ou mais como multidimensionais.

Duas definições de dimensionalidade são encontradas na literatura: unidimensionalidade restrita e unidimensionalidade essencial (STOUT, 1990; JUNKER, 1993; WEI, 2008).

Unidimensionalidade restrita implica que as probabilidades de respostas corretas para os itens do teste são restritamente função de somente um traço latente, além das variáveis que representam as características do item. A fórmula geral para o modelo da TRI com unidimensionalidade restrita pode ser representada pela equação (2). Na prática, segundo Wei (2008) a unidimensionalidade restrita é considerada demasiada restritiva para ser aplicável em situações do mundo real, onde as dimensões menores do traço latente a ser medido também afetam as respostas do examinando.

Unidimensionalidade essencial é uma definição menos rigorosa e reconhece que um teste é intrinsecamente multidimensional, mas as respostas dos itens são afetadas por um traço latente dominante e alguns fatores latentes não-significantes. Segundo Wei (2008), nesta definição assume-se um vetor de traços latentes  $\theta_j = \{\theta, \theta_1, \theta_2, \dots, \theta_M\}$  relevantes para o teste, onde  $\theta$  representa o traço latente dominante e o restante do vetor denota os traços latentes menos relevantes, mas que estão associados ao teste. Um modelo da TRI com unidimensionalidade essencial pode ser representada pela equação (6).

$$P(U_{ij} = u_{ij} | \theta_j = \theta) = f(\theta, b_i, u_{ij}) \quad (6)$$

Ainda segundo Wei (2008), a definição de unidimensionalidade restrita é impraticável para testes psicológicos e pedagógicos, mas a unidimensionalidade essencial apresenta uma adequada aproximação. Atender apenas o traço latente dominante e ignorar outras características não essenciais ou menores não é prejudicial para quaisquer efeitos práticos e não tem impacto negativo sobre as estimativas dos examinados.

Como citado anteriormente, modelos multidimensionais da TRI podem ser extensões de modelos unidimensionais da TRI. Ao invés de manter o pressuposto de unidimensionalidade restrita ou essencial os modelos multidimensionais assumem que há mais de uma dimensão latente que influencia significativamente as respostas dos examinandos. Para Reise et al., (2007), na maioria dos casos, a multidimensionalidade é devido a

heterogeneidade do conteúdo do item que é necessário para representar adequadamente a complexidade do traço latente.

Segundo Wei (2008), do ponto de vista matemático, a unidimensionalidade essencial é tecnicamente um modelo multidimensional, mas se aproxima de um modelo unidimensional no limite. Em outras palavras, é multidimensional por natureza, mas as previsões para as respostas aos itens são feitas com base na aproximação da abordagem unidimensional. Para mais detalhes matemáticos da comparação entre modelos da TRI unidimensional essencial e multidimensional ver Nandakumar (1991) e Wei (2008).

Dickenson (2005) mostra que, quando um modelo unidimensional da TRI é usado na análise de dados multidimensional, a dimensão resultante única é, na verdade, um composto linear de múltiplas dimensões. Segundo Ackerman (1994), a interação entre um grupo de examinados a um item pode ser modelado unidimensionalmente por um item que pode medir apenas uma dimensão ou um composto de múltiplas dimensões. Entretanto, segundo Ackerman (1994), quando a interação entre um grupo de examinados e vários itens é considerada, a suposição de unidimensionalidade deve ser considerada com muito cuidado. Por exemplo, um exame de matemática pode ser desenvolvido para medir a capacidade de álgebra, geometria, probabilidade, medidas, número e operações. Embora estas capacidades possam estar relacionadas, a relação dificilmente será perfeita. Por exemplo, dois examinandos com igual habilidade em geometria pode ter diferentes habilidades em álgebra (LIN, 2008).

Min (2003) resume três diferentes condições sob as quais a aplicação dos modelos unidimensionais é adequada: (a) a capacidade do examinado e as características do item variam em uma dimensão, como presume o modelo, (b) a capacidade do examinado varia apenas em uma dimensão mesmo que os itens do teste estejam medindo mais de uma habilidade, (c) a habilidade dos examinados é diferente dentro de múltiplas dimensões, mas todos os itens estão medindo o mesmo composto de habilidades. Segundo Lin (2008), em outras condições que não podem ser classificadas dentro destas três condições, aplicar modelos unidimensionais pode ser problemático. Estudos têm mostrado que quando os dados multidimensionais são modelados com base no pressuposto unidimensional, erro de medida irá aumentar e as consequências dos resultados são problemáticas (ANSLEY; FORSYTH, 1985; SIRECI, et al., 1991; ACKERMAN, 1994; RECKASE, 1995).

A dimensionalidade de um teste pode ser avaliada por uma variedade de índices de ajuste e testes de significância e quase todos são baseados no conceito de unidimensionalidade essencial (TATE, 2003).

Segundo Wei (2008), o princípio básico para avaliação da dimensionalidade de um teste segue mais ou menos a seguinte lógica: a suposição de independência local é avaliada sob um modelo hipotetizado, unidimensional ou multidimensional, e medidas são obtidas para indicar a quantidade de dependência entre os itens. Se a dependência entre os itens é mais forte do que seria esperado por acaso, a suposição de independência local é suspeita. Se existir forte dependência entre os itens, a dimensionalidade assumida do modelo será rejeitada. Nesta abordagem, a independência local é definida como homogênea ou fraca e as dependências dos itens são calculadas com base nas covariâncias condicionais das respostas do item para todos os pares de itens e em todos os níveis de traços latentes. Neste caso, quando as condições de unidimensionalidade são satisfeitas para um dado teste e uma específica população, diferenças no desempenho do teste são atribuídas às diferenças dos indivíduos.

Estes desenvolvimentos gerais de análise de fatores costumam estar associados às análises de variáveis contínuas. Como o presente trabalho trata de respostas dicotômicas, algumas aproximações são necessárias, estas geram métodos particulares como, por exemplo, métodos de informação restrita e os métodos de informação plena.

Segundo Soares (2005), o método de informação restrita consiste na inspeção dos valores próprios da matriz de correlação tetracórica quanto aos demais. Detalhes sobre matriz de correlação podem ser encontrados em Divgi (1979). Este método está disponível nos softwares Testfact 2.13 (WILSON, et al., 1998) e R, pacote *psych* (R CORE TEAM 2012). Com base neste método, Franklin et al. (1995) e Ledesma e Valero-Mora (2007), sugerem a análise paralela, que é uma técnica de simulação de Monte Carlo a qual segue a lógica geral defendida por Wei (2008), ou seja, primeiramente é realizada uma análise de dimensionalidade dos dados reais, com base na avaliação dos valores próprios (matriz de correlação tetracórica). Após isto, é gerado um conjunto de dados simulados, que não tem qualquer relação entre os itens, mas que tem a mesma proporção de respostas corretas dos dados reais e o mesmo tamanho da amostra. Os dados gerados são então analisados da mesma forma que os dados reais. Uma vez que os dados gerados têm a mesma distribuição dos dados reais, a análise destes dados tem a mesma tendência em relação a dificuldade dos itens reais. Os valores próprios para os dados gerados são então representados no mesmo gráfico com os correspondentes dados reais e, o número de valores próprios para os dados reais que estiverem acima dos apresentados para os

dados simulados representa o número de dimensões do conjunto de dados reais. Exemplos práticos desta abordagem podem ser encontradas em Reckase (2009), Townsend e Konold (2010) entre outros.

Os métodos de análise fatorial da informação plena (BOCK; AITKIN, 1981; BOCK, et al., 1988; MURAKI; CARLSON, 1995; MURAKI; ENGELHARD, 1985) foram desenvolvidos a partir de uma adaptação do modelo tradicional de análise fatorial (JOHNSON; WICHERN, 2007) que considera a estrutura de dimensões associadas a variáveis contínuas. Este método está disponível nos softwares Testfact 2.13 (WILSON, et al, 1998), R, pacote *mirt* (R CORE TEAM 2012) e flexMIRT<sup>TM</sup> (CAI, 2012).

Quando o pressuposto da unidimensionalidade não é viável para um teste, um modelo multidimensional deve ser considerado. Modelos multidimensionais da TRI são aplicados tanto para usos exploratório como confirmatório. Modelos exploratórios da TRIM são usados quando o desenvolvedor de teste não tem uma teoria sobre a estrutura do teste a ser analisado. Quando o desenvolvedor possui um conhecimento aprofundado sobre os domínios de conteúdo ou habilidades cognitivas avaliadas no teste e como eles são combinados para uma resposta correta a cada item, modelos confirmatórios são mais indicados (WEI, 2008).

Cada um dos métodos para avaliação de dimensionalidade baseadas na análise de fatores, como por exemplo, a razão do primeiro e do segundo autovalor, avaliação do *scree plot*, inspeção da distribuição do resíduo depois da extração de um fator, análise confirmatória ou métodos mais sofisticados como análise paralela modificada (CHERNYSHENKO et al., 2001) ou, ainda, métodos não paramétricos, possuem seus méritos e são úteis para informar a quantidade e/ou predominância de um fator. Entretanto, segundo Reise et al., (2007), estas abordagens não são altamente informativas em termos de abordar as questões-chave na modelagem da TRI, tais como: informar especificamente qual o grau de violação da unidimensionalidade que afeta a modelagem dos dados na utilização de modelos unidimensionais da TRI, e qual modelo da TRIM deve ser usado. Como alternativa Reise et al. (2007) sugerem a utilização mais sistemática de modelos hierárquicos, em particular, o modelo bi-fatorial (GIBBONS; HEDEKER, 1992; GIBBONS et al., 2007). Em uma representação bi-fatorial cada item da escala pode ter uma carga positiva em um traço latente geral que se presume ser a base de todos os itens. Normalmente, esse fator geral vai ser conceitualmente mais amplo, e é a característica que o pesquisador está mais interessado em medir. Além disso, cada item pode carregar em um ou mais fatores. Esses fatores tenderão a ser conceitualmente mais estreito. Segundo Chen et al., (2006), o modelo bi-

fatorial é potencialmente aplicado quando (a) existe um fator geral que é a hipótese para explicar a comunalidade dos itens; (b) existem múltiplos fatores específicos e cada um é hipoteticamente influenciador de um fator geral, e (c) o pesquisador pode estar interessado em fatores específicos, bem como no fator dominante.

De uma forma geral, a seleção de um modelo da TRI adequado é baseado em parte na adequação do modelo aos dados. Se os dados não se enquadram no modelo, as propriedades do modelo da TRI, tais como a invariância dos parâmetros para a população, não será assegurada (HAMBLETON et al., 1991).

Na próxima sessão serão apresentados alguns modelos multidimensionais da TRI e suas peculiaridades.

### 2.3.3 Modelos Multidimensionais da TRI

Dependendo dos objetivos finais e da estrutura dos dados, a TRIM pode ser considerada como um caso especial de análise estatística multivariada, especialmente análise fatorial ou modelagem de equações estruturais, ou ainda como uma extensão da TRI unidimensional (RECKASE, 2009).

No primeiro caso, a TRIM pode ser considerada um tipo de análise fatorial (AF) para dados observados não contínuos (dicotômicos ou politômicos) (FRAGOSO, 2008). Segundo Reckase (1997), em alguns casos a formulação estatística da AF é virtualmente idêntica a da TRIM, tanto que softwares que utilizam a metodologia de análise fatorial de informação plena apresentada por Bock et al. (1988) podem ser usados tanto para análise fatorial como para TRIM. O que diferencia as duas metodologias é essencialmente a interpretação, ou seja, a AF exploratória, historicamente, tem sido usada para encontrar o menor número possível de variáveis hipotéticas ou fatores subjacentes que explicam as relações dentro de um grande número de variáveis empíricas. Isto é, a AF exploratória é basicamente uma técnica de redução de dados (RECKASE, 1997). A TRIM, entretanto, foca na modelagem da interação entre respondentes e itens. Desta forma, segundo Reckase (2009), a TRIM difere da AF nos casos em que as diferentes características das variáveis de entrada (itens), como o grau de dificuldade e a discriminação, são consideradas importantes. Além disso, a TRI possui a possibilidade de modelar as respostas ao acaso, o que não é possível em modelos tradicionais de estatística multivariada.

As primeiras contribuições teóricas na área de análise fatorial que impulsionaram o desenvolvimento da TRIM foram Horst (1965), McDonald (1967), Christoffersson (1975), Muthén (1978), Bock e Aitken (1981), entre outros. Estes autores propuseram modificações nos modelos de AF com o objetivo de modelar também as variáveis de entrada, modelando a probabilidade de uma resposta correta a um conjunto de itens multidimensionais (MUTHÉN 1978), além de modelagem da AF não linear (McDONALD 1967) e inserção do modelo da ogiva normal neste contexto (BOCK; AITKEN, 1981). Mais detalhes sobre o histórico da TRIM podem ser encontrados em McDonald (2000) ou Reckase (1997, 2009).

A escolha de um modelo multidimensional específico depende, principalmente, da natureza das dimensões latentes, dos itens e da forma como estes se associam dentro do teste. Diferentes modelos da TRIM assumem diferentes relações estatísticas entre as dimensões latentes e o desempenho do respondente (HARTIG; HOHLER, 2009). Além disso, antes de apresentar os modelos da TRIM é importante compreender as estruturas multidimensionais envolvidas em um teste. Primeiramente, a relação entre dimensões latentes e os itens; e na sequência, a relação entre as dimensões latentes e os respondentes. Segundo Adams et al. (1997) e Hartig e Hohler (2009), o padrão de relações entre as dimensões e os itens pode ser definido por uma matriz de carga com uma estrutura simples (multidimensionalidade entre os itens) ou por uma estrutura de carregamento complexo (multidimensionalidade dentro de cada item) e, portanto, varia em sua complexidade. Já o padrão de relação entre as dimensões latentes e a habilidade do respondente possui uma interação compensatória ou não compensatória.

Multidimensionalidade entre itens: representa testes que contenham várias sub-escalas relacionadas, mas supostamente em distintas dimensões latentes. Em tais testes, cada item pertence a uma única sub-escala particular e não há itens em comum entre as sub-escalas (WILSON; HOSKENS, 2005). No passado a modelagem destes itens era feita por meio da aplicação de um modelo unidimensional para cada uma das sub-escalas, conhecidas como abordagens consecutivas (DAVEY; HIRSCH, 1991), ou então, simplesmente ignoravam a multidimensionalidade e tratavam o teste como unidimensional. Nesta última abordagem, não é possível encontrar a relação entre as dimensões e ela não é adequada quando as dimensões não são fortemente correlacionadas. Na abordagem consecutiva existem problemas relacionados à mensuração do erro em cada dimensão e com a incapacidade de utilizar todos os dados disponíveis (WANG, 1994). Uma alternativa para equacionar estes problemas, dentro desta natureza de itens,

é o MRCMLM (*Multidimensional Random Coefficients Multinomial Logit Model*) proposto por Adams et al. (1997). A vantagem de um modelo como o MRCMLM com dados deste tipo é que: (1) reconhece explicitamente a estrutura pretendida do teste, (2) fornece estimativas diretas das relações entre as dimensões latentes e, (3) baseia-se a relação (por vezes forte) entre as dimensões latentes para produzir estimativas mais precisas de parâmetros e medições individuais. Mais detalhes do MRCMLM é apresentado mais a frente.

Multidimensionalidade dentro de cada item: ocorre quando o conjunto de itens do teste mede mais de uma dimensão latente e alguns itens requerem habilidade em mais de uma dimensão para serem respondidos (WILSON; HOSKENS, 2005). Segundo Wei (2008), a multidimensionalidade dentro de cada item é frequentemente analisada com modelos compensatório ou não-compensatórios, dependendo da natureza das interações entre as dimensões. Modelos compensatórios são usados mais frequentemente que os não-compensatórios em situações práticas porque apresentam maior facilidade de estimação dos parâmetros. Para Hartig e Hohler (2009) modelos que incorporam multidimensionalidade dentro de cada item são adequados para modelar interações entre as dimensões latentes e os itens. Aqui, a probabilidade de resolver um item pode ser modelada como uma função de uma combinação de diferentes dimensões latentes. Assim, dimensionalidade dentro de cada item implica suposições explícitas sobre as competências requeridas para os diferentes itens, o que exige fortes pressupostos teóricos. Dentre estes pressupostos está a necessidade de saber a relação entre as dimensões latentes e os indivíduos, ou seja, se estas são compensatórias ou não-compensatórias (HARTIG; HOHLER 2009). Consequentemente, os modelos com multidimensionalidade dentro de cada item, são particularmente interessantes para a modelagem de traços latentes complexos que não podem ser explicados por uma única dimensão.

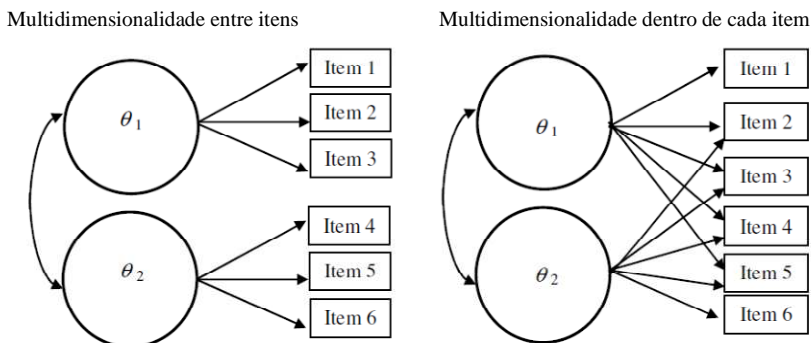
Segundo Hartig e Hohler (2009), a vantagem de modelos que utilizam a estrutura multidimensional entre itens é que estes são menos complexos que os modelos que utilizam a multidimensionalidade dentro de cada item, e o traço latente pode ser facilmente interpretado. Dentro destes modelos, a estimação dos escores do traço latente fornece uma medida simples do desempenho em um conjunto específico de itens. Em muitos casos, estas medidas serão altamente correlacionadas, pois os itens medem o mesmo conjunto de habilidades. Entretanto, as dimensões latentes no modelo multidimensional entre itens representam a combinação necessária de todas as habilidades requeridas para resolver os respectivos itens, independentemente de como essas habilidades precisam ser integradas.



Qualquer sobreposição é representada nas correlações latentes. Assim, se o interesse principal da pesquisa é obter medidas descritivas do desempenho em áreas de determinados conteúdos, os modelos entre itens são mais adequados do que modelos mais complexos como os de multidimensionalidade dentro de cada item (HARTIG; HOHLER 2009).

A distinção entre a multidimensionalidade dentro de cada item e entre itens é ilustrada na Figura 5.

Figura 5 - Distinção entre a multidimensionalidade dentro de cada item e entre itens.



Embora o tipo de dimensionalidade seja, conceitualmente, um conceito básico, a distinção entre multidimensionalidade entre itens e dentro de cada item é vital para a correta identificação dos modelos TRIM (BABCOCK, 2009).

McDonald (2000) argumenta que uma condição suficiente para identificar corretamente um modelo multidimensional compensatório com fatores correlacionados é que cada dimensão latente possua pelo menos dois itens que carreguem apenas naquela única dimensão. Para fatores ortogonais, a identificação adequada requer três itens.

Nos modelos com a multidimensionalidade dentro de cada item, as múltiplas dimensões que são necessárias para itens individuais podem ser integradas de diferentes maneiras, dependendo da natureza da relação entre as dimensões latentes e os respondentes. Deste ponto de vista, os modelos da TRIM geralmente são classificados como compensatórios ou não-compensatórios. de Ayala (2009) dá dois exemplos que descrevem de forma prática a diferença entre estes modelos: primeiro, considerando um instrumento projetado para medir auto-eficácia em superar barreiras para

uma alimentação saudável, teoricamente, este traço latente envolve uma dimensão cognitiva e outra afetiva, desta forma, a resposta aos itens deste instrumento são função da localização dos respondentes nestas dimensões. O segundo exemplo é o desempenho em uma prova de matemática, neste caso, tem-se, por exemplo, duas dimensões, uma delas reflete a proficiência em matemática e a outra a proficiências em leitura. No segundo exemplo, um indivíduo com alta proficiência em leitura pode compensar uma possível baixa proficiência em matemática. Em contraponto, no primeiro exemplo a localização de um respondente na dimensão cognitiva não pode compensar sua localização na dimensão afetiva. Desta forma, o problema de matemática representa o modelo compensatório e o instrumento para medir auto-eficácia representa um exemplo de modelo não-compensatório.

Reckase (1985) descreve o modelo multidimensional compensatório de dois parâmetros da seguinte forma:

$$P(U_{ij} = 1 | \theta'_{jk}, d_i, a'_{ik}) = \frac{e^{(\sum_{k=1}^m a_{ik}\theta_{ik} + d_i)}}{1 + e^{(\sum_{k=1}^m a_{ik}\theta_{ik} + d_i)}}, \quad (7)$$

onde:  $U_{ij}$  = resposta da pessoa  $j$  ao item  $i$  (0 ou 1);

$a_{ik}$  = vetor do parâmetro de discriminação do item  $i$  na dimensão  $k$ ;

$\theta_{jk}$  = vetor do traço latente da pessoa  $j$  na dimensão  $m$ ;

$d_i$  = escalar do parâmetro de dificuldade do item  $i$ .

O expoente de  $e$  no modelo (7) pode ser escrito da seguinte forma:

$$\sum_{k=1}^m a_{ik}\theta_{ik} + d_i = a_{i1}\theta_{j1} + a_{i2}\theta_{j2} + \dots + a_{ik}\theta_{jk} + d_i \quad (8)$$

A equação (8) mostra que o expoente é uma função linear de elementos de  $\theta$  com o parâmetro  $d$ , como a ordenada na origem e os elementos do vetor  $a$  como os parâmetros de inclinação/discriminação. Uma das propriedades deste modelo é que a expressão representada no expoente define uma reta num espaço de  $k$  dimensões que pode gerar linhas de equiprobabilidade, ou seja, esta forma multidimensional permite que existam infinitas combinações lineares que resultam no mesmo expoente,

gerando, portanto, mesma probabilidade de acerto. Esta propriedade confere ao modelo a característica compensatória.

Nota-se, ainda, que neste modelo, para cada item, existe um parâmetro de discriminação ( $a$ ) e um parâmetro de habilidade ( $\theta$ ) para cada dimensão e apenas um parâmetro de dificuldade para todas as dimensões. Esta particularização do modelo gera algumas interpretações vetoriais destes parâmetros. Por exemplo, segundo Reckase (1997), o número de dimensões utilizadas para representar a interação item – sujeito vai depender do objetivo da análise. Estas dimensões podem não ter significado prático, e se possuem ou não é uma questão de validação empírica do construto. Embora o espaço multidimensional possa sofrer rotações de forma a alinhar-se o eixo de  $\theta$  com pontos com significado no espaço, estas rotações podem ou não reter a estrutura da covariância inicial das dimensões de  $\theta$ . O modelo não obriga que as rotações sejam ortogonais, podendo, portanto, ser rotacionada obliquamente ou ortogonalmente.

Os parâmetros de discriminação do modelo são dados pelos elementos do vetor  $a_{ik}$ , estes podem ser interpretados da mesma maneira que o parâmetro de discriminação do modelo unidimensional representado pela equação (4). Sendo assim, os elementos deste vetor estão relacionados com a inclinação da superfície de resposta do item na direção do correspondente eixo  $\theta$  (RECKASE 1997). Por se tratar de um conjunto vetorial, pode-se representar este conjunto por um vetor geral que representa a soma de todos os vetores, chamado de discriminação multidimensional do item  $i$  (MDISC), a equação deste parâmetro pode ser representada da seguinte maneira (RECKASE 1986):

$$A_i = \sqrt{\sum_{k=1}^m a_{ik}^2} \quad (9)$$

$A_i$  corresponde ao comprimento de vetor do item  $i$  e representa a quantidade máxima de discriminação.

O parâmetro  $d_i$ , neste modelo, encontra-se relacionado com a dificuldade do item. O valor deste parâmetro não pode ser interpretado da mesma forma que o valor do parâmetro  $b$  dos modelos unidimensionais da equação (4) (RECKASE, 1997).

A forma usual para representar o expoente do modelo unidimensional é:

$$a(\theta - b) = a\theta - ab \quad (10)$$

onde,  $d_i = -a_i * b_i$ . Um valor (no modelo multidimensional) que é equivalente à interpretação do parâmetro  $b$  no modelo unidimensional é dado por (RECKASE, 1985):

$$B_i = \frac{-d_i}{\sqrt{\sum_{k=1}^m a_{ik}^2}} \quad (11)$$

Esta medida indica a dificuldade de um item (MDIFF), corresponde a localização do vetor no espaço (direção e intensidade).  $B_i$  positivo indica item  $i$  difíceis  $B_i$  negativo indica itens fáceis. A distância da origem até ao ponto de inclinação mais acentuada do vetor  $A_i$  indica o nível de dificuldade do item. Esta interpretação de  $B$  aplica-se apenas à direção especificada pelo vetor  $A_i$  (de Ayala, 2009).

A localização multidimensional do item corresponde a direção angular de cada item em relação ao eixo de dimensão  $k$ . E pode ser representado matematicamente por (12) (RECKASE, 1985).

$$\cos \alpha_{ik} = \frac{a_{ik}}{\sqrt{\sum_{k=1}^m a_{ik}^2}} \quad (12)$$

onde,  $\alpha_{ik}$  é o ângulo que a linha a partir da origem do espaço até ao ponto da inclinação mais acentuada faz com o eixo  $k$ .

O modelo de probabilidade para construtos multidimensionais não-compensatórios e de dois parâmetros foi proposto por Simpson (1978) e pode ser escrito da seguinte forma:

$$P(X_{ij} = 1 | \theta'_{jk}, b'_i, a'_i) = \prod_{k=1}^m \frac{e^{a_{ik}(\theta_{jk} - b_{ik})}}{1 + e^{a_{ik}(\theta_{jk} - b_{ik})}} \quad (13)$$

Neste modelo, ao invés de se ter um escalar  $d_i$  para representar a dificuldade do item, tem-se vetor  $b_i$  para representar a dificuldade do item  $i$  na dimensão  $k$ . Sendo assim, neste modelo, cada item possui um parâmetro de discriminação e um parâmetro de dificuldade para cada dimensão. A multiplicação natural dos componentes impede que um respondente com baixa proficiência em uma dimensão compense sua proficiência total com outra dimensão (ACKERMAN, 1994).

No caso dos modelos multidimensionais de três parâmetros, tanto no caso compensatório como no caso não-compensatório é inserido o parâmetro de acerto casual que é o mesmo para todas as dimensões, desta forma, adaptando a equação (7) para transformá-la em uma equação multidimensional compensatória de três parâmetros, tem-se:

$$P(X_{ij} = 1|\theta_{jk}, d_i, c_i, a_{ik}) = c_i + (1 - c_i) \frac{e^{(\sum_{k=1}^m a_{ik}\theta_{jk} + d_i)}}{1 + e^{(\sum_{k=1}^m a_{ik}\theta_{jk} + d_i)}} \quad (14)$$

onde  $c_i$  representa o parâmetro de acerto casual.

No caso do modelo compensatório, adaptando-se a equação (13), tem-se:

$$P(X_{ij} = 1|\theta_{jk}, b_{ik}, c_i, a_{ik}) = c_i + (1 - c_i) \prod_{k=1}^m \frac{e^{a_{ik}(\theta_{jk} - b_{ik})}}{1 + e^{a_{ik}(\theta_{jk} - b_{ik})}} \quad (15)$$

Seguindo a mesma linha de raciocínio, outros autores propuseram extensões dos modelos unidimensionais da TRI para respostas politômicas, tais como Modelo de Crédito Parcial Multidimensional (KELDERMAN; RIJKES, 1994); Modelo de Crédito Parcial Generalizado Multidimensional (YAO; SCHWARZ, 2006), Modelo de Resposta Gradual Multidimensional (MURAKI; CARLSON, 1995). Segundo Babcock (2009), a maioria destes modelos multidimensionais politômicos considera como base o modelo compensatório.

A generalização multidimensional do modelo de Rasch é mais complexa do que a generalização dos modelos de dois e três parâmetros, particularmente porque o modelo de Rasch considera o parâmetro de discriminação igual a 1 para todos os itens e, considerando que no modelo compensatório o parâmetro de discriminação é o único que varia entre as dimensões, a limitação matemática tornaria o modelo um simples modelo unidimensional. Para mais detalhes matemáticos desta particularidade ver

Reckase (2009) e para casos não-compensatórios utilizando generalização multidimensional do modelo de Rasch ver Rijmen e de Boeck (2005).

Para equacionar o problema na generalização do modelo de Rasch em casos compensatórios Adam et al (1997) propuseram o modelo *Multidimensional Random Coefficients Multinomial Logit Model* – MRCMLM que é apresentado na equação (16) como uma extensão multidimensional do modelo de Rasch apresentado na equação (3)

$$P(U_{ik} = 1|A, B, \xi, \theta) = \frac{e^{b_{ik}\theta + a'_{ik}\xi}}{\sum_{k=0}^{K_i} e^{b_{ik}\theta + a'_{ik}\xi}} \quad (16)$$

onde,  $A$  é a matriz com os elementos  $a_{ik}$ ,  $B$  é a matriz com os elementos  $b_{ik}$ , que indica a dimensão ou dimensões que são requeridas para obter o escore de  $k$  no item,  $\xi$  é o vetor do parâmetro de dificuldade do item, e  $\theta$  é um vetor das coordenadas da posição de uma pessoa. A equação (16) apresenta uma generalização deste modelo para dados politômicos. Segundo Reckase (2009) para o caso de dados dicotômicos, usa-se a mesma notação apenas limitando o valor de  $k$  entre 0 e 1, conforme equação (17).

$$P(U_{ij} = 1|a_i, d_i, \theta_j) = \frac{e^{a_i\theta'_j + d_i}}{1 + e^{a_i\theta'_j + d_i}} \quad (17)$$

Onde,  $a_i$  é um vetor em que  $a_i = b_{ik}$  e  $d_i$  é um escalar igual a  $a'_{ik} \cdot \xi$ . Nota-se que quando  $k = 0$  na equação (16) o expoente de  $e$  é zero então o primeiro termo da soma no denominador fica igual a 1. A equação (17) é aparentemente idêntica a equação (7) se considerarmos o somatório das dimensões. Entretanto, a diferença entre os dois está na maneira em que o vetor  $a_i$  é especificado. Na equação (7),  $a_i$  é uma característica do item  $i$  que é estimada enquanto que em (17),  $a_i$  é uma característica do item  $i$  que é especificada pelo desenvolvedor do teste – mesma lógica do modelo de Rasch (RECKASE, 2009). Adams et al. (1997) especificaram o MRCMLM para os dois casos de dimensionalidade – entre itens e dentro de cada item. Para dimensionalidade entre itens o vetor  $a_i$  possui valor zero para todas as dimensões, exceto para uma que especifica a dimensão que é medida pelo item. Isto é, o desenvolvedor do teste especifica a dimensão do item, e o processo de estimação é feito considerando esta limitação. Já no caso da dimensionalidade dentro de cada item, o vetor  $a_i$  tem valor diferente de zero em mais de uma dimensão e o ajuste dos valores estimados também dependem da definição do desenvolvedor. Mais detalhes ver Adams et al. (1997).

O MRCMLM foi aplicado na prova de matemática do PISA 2003 (OECD, 2005). Como o MRCMLM não permite relações hierárquicas entre as dimensões (HARTIG; HOHLER, 2009) o PISA 2003 foi avaliado com 7 dimensões considerando a multidimensionalidade entre itens.

A estrutura multidimensional dentro de cada item foi aplicada por Walker e Berebts (2003) também na área educacional, e Hartig e Höhler (2008) que fazem uma aplicação bidimensional na avaliação de competências em testes de línguas para estrangeiros, entretanto, estes impõem algumas restrições além das impostas pelo próprio modelo de Rasch (mesmo parâmetro de discriminação para todos os itens) que é o mesmo vetor de parâmetros dos itens para todo o teste.

Gibbons e Hedeker (1992) generalizaram o trabalho clássico de Holzinger e Swineford (1937) e propuseram o modelo *Full-information Bi-fatorial* (FI Bi-fatorial) para dados dicotômicos (LI; RUPP, 2011). Este modelo consiste em um fator geral e grupos de fatores ou dimensões independentes. O modelo FI Bi-fatorial assume a presença de um fator geral que envolve todos os itens e dois ou mais grupos de fatores (ou dimensões) correspondentes de subgrupos específicos (GIBBONS et al., 2007).

Matematicamente, o modelo FI Bi-fatorial considera casos em que para  $n$  itens, existe uma solução com  $s$  fatores dos quais, um é considerado um fator geral e  $s - 1$  grupos ou fatores relacionados. A solução bi-fatorial restringe cada item  $i$  a ter uma carga diferente de zero  $a_{i1}$  sobre a dimensão primária e uma segunda carga ( $a_{ik}$ ,  $k = 2, \dots, s$ ) e não mais do que um dos  $s - 1$  grupos de fatores. Para quatro hipotéticos itens, a matrix padrão bi-fatorial pode ser representada da seguinte forma (GIBBONS; HEDEKER 1992):

$$\alpha = \begin{bmatrix} \alpha_{11} & \alpha_{12} & 0 \\ \alpha_{21} & \alpha_{22} & 0 \\ \alpha_{31} & 0 & \alpha_{33} \\ \alpha_{41} & 0 & \alpha_{43} \end{bmatrix} \quad (18)$$

onde a primeira coluna da matriz representa o fator principal, e a segunda e terceira colunas representam o grupo de fatores específicos.

Segundo Seo (2011), a estrutura dimensional em um modelo bi-fatorial é pré-determinada, através de informação prévia. Portanto, o modelo bi-fatorial é um modelo confirmatório. Na perspectiva de uma abordagem confirmatória, o modelo permite que cada item tenha cargas em um único fator geral, e apenas um fator específico do grupo. Esta particularidade reduz o número de parâmetros a serem estimados e dá o modelo mais graus de liberdade. Além disso, o modelo bi-fatorial pode

evitar o problema de estimar correlações inter fatores, porque o fator geral contribui diretamente para todos os itens, e os fatores secundários para a informação residual remanescente após o cálculo do fator geral são independentes uns dos outros. Uma particularização deste modelo é o fato de que os fatores secundários são necessariamente ortogonais entre si e em relação ao fator geral (GIBBONS, HEDEKER 1992).

Para dados binários, o modelo bi-fatorial pode ser definido como uma particularização do modelo multidimensional compensatório apresentado na equação (7). No caso do modelo bi-fatorial é inserida a restrição de carregamento dos parâmetros de discriminação como pode ser visto na equação (19).

$$P(U_{ij} = 1 | \theta_{jg}, \theta_{jesk}) = \frac{e^{[(a_{ig}\theta_{jg} + a_{iesk}\theta_{jesk}) - d_i]}}{1 + e^{[(a_{ig}\theta_{jg} + a_{iesk}\theta_{jesk}) - d_i]}} \quad (19)$$

Onde  $\theta_{jg}$  é a habilidade do respondente  $j$  no fator geral,  $\theta_{jesk}$  é a habilidade do respondente  $j$  no fator específico  $k$ ,  $a_{ig}$  representa o parâmetro de discriminação do item  $i$  no fator geral e  $a_{iesk}$  representa o parâmetro de discriminação do item  $i$  no fator específico  $k$  e por fim,  $d_i$  representa o escalar do parâmetro de dificuldade do item  $i$  referente a dimensão geral e a dimensão específica  $k$ . Neste modelo, assim como no modelo multidimensional compensatório representado na equação (7) as respostas são assumidas como sendo estatisticamente independentes.

Segundo Gibbons et al. (2007), o modelo FI Bi-fatorial é relevante sempre que os itens compartilham uma característica comum. A presença dos subgrupos de itens tipicamente introduz associações relevantes ao teste que não podem ser verificadas com atribuição total das cargas ao fator geral. Além disso, segundo os autores, esta separação de fatores melhora o erro das estimativas.

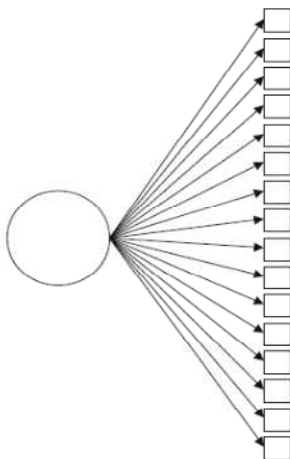
Gibbons e Hedeker (1992) e Gibbons et al. (2009) argumentam que as restrições do modelo bi-fatorial apresentadas na matrix (18) conduzem a uma maior simplificação das equações de probabilidade por requerer somente a avaliação de duas dimensões integrais o que (a) permite análises de modelos com grande número de fatores (ou dimensões), (b) permite dependência condicional entre subgrupos de itens identificados, e (c) em muitos casos fornece uma solução mais parcimoniosa do que a análise fatorial de informação plena, devido a sua falta de restrições. Gibbons et al. (2007b) estendeu o modelo bi-fatorial para casos de itens politômicos.

A Figura 6 contextualiza o modelo FI Bi-fatorial dentro de algumas estruturas multidimensionais.

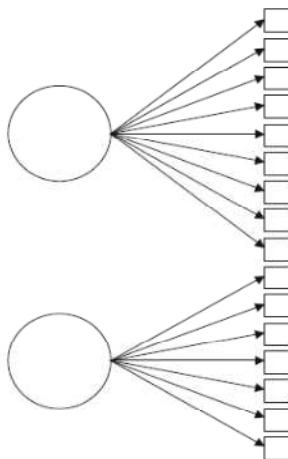


Figura 6 - Quatro possíveis estruturas de traços latentes, Reise et al. (2007).

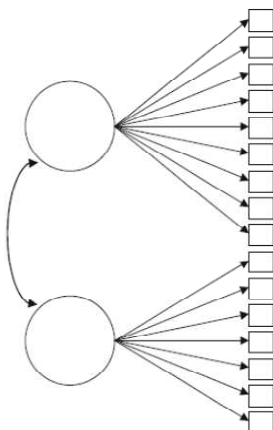
Modelo A



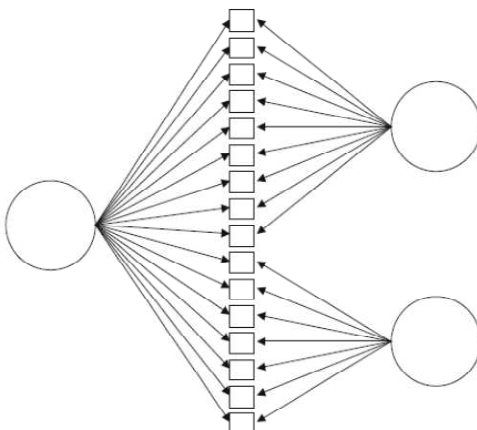
Modelo B



Modelo C



Modelo D



O modelo A é o modelo padrão unidimensional no qual a covariância entre as respostas aos itens é explicada por um fator comum. No modelo B, a matriz de dados contém mais de uma dimensão comum,

porém, as dimensões não são correlacionadas. Este é um caso trivial de multidimensionalidade e é de fácil solução, formando subescalas e, em seguida, ajustar o modelos unidimensionais da TRI para cada sub-escalas em separado. Esta é essencialmente equivalente a assumir as dimensões como não correlacionadas.

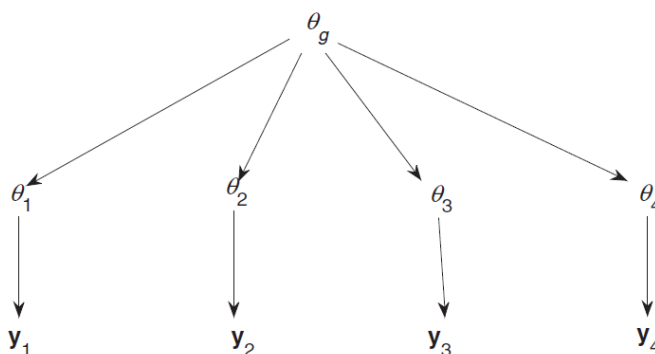
Modelo C também tem mais de um fator comum entre os itens, porém os fatores são correlacionados. Tal representação caracteriza-se como um modelo multidimensional não hierárquico.

Finalmente, o modelo D é um modelo bi-fatorial, ou seja, existe um fator geral, que explica as correlações entre item, mas, além disso, há também os chamados fatores de "grupo" (no lado direito da figura) que tentam capturar a covariância dos itens que é independente da covariância do fator geral. Em outras palavras, expressa em termos de uma medida de qualidade de websites, pode-se pré supor que o traço latente qualidade, tendo em vista sua amplitude conceitual, representa um fator geral explicado por outros fatores (por exemplo, usabilidade, estética, arquitetura da informação, etc), caracterizando então uma adequação ao modelo D (multidimensional hierárquico), entretanto a amplitude deste conceito pode dar lugar a dissolução deste em subfatores correlacionados – Modelo C (multidimensional não hierárquico).

Comparando-se os modelos da Figura 5 com os modelos da figura 6 verifica-se que o modelo FI bi-fatorial é um modelo de carregamento complexo, ou seja, de multidimensionalidade dentro de cada item, e com restrições que o tornam também um modelo de dimensionalidade entre itens hierárquicos. Outros modelos surgiram com base no FI bi-fatorial, como por exemplo, o modelo Testlet (BRADLOW et al., 1999) que é um caso especial do modelo FI bi-fatorial. É obtida restringindo os carregamentos dentro das dimensões específicas para serem proporcionais à carga geral (LI et al., 2006; RIJMEN, 2010).

O modelo de segunda ordem da TRIM, também é uma particularidade do modelo Bi-fatorial (RIJMEN, 2010). Neste caso, existe um fator específico para cada dimensão, assim como o modelo Bi-fatorial e o modelo Testlet, e um fator geral, mas, ao contrário destes modelos, os itens não dependem diretamente deste fator geral. Os itens dependem diretamente das respectivas dimensões específicas, que por sua vez dependem da dimensão geral, como mostra a figura 7.

Figura 7 - Estrutura do modelo de segunda ordem da TRIM, Rijmen, 2010.



A Figura 7 mostra a estrutura hierárquica modelada pelo modelo de segunda ordem da TRIM, onde  $\theta_g$  representa a dimensão geral  $\theta_1, \theta_2, \theta_3, \theta_4$  representam as dimensões específicas e  $y_1 y_2 y_3 y_4$  representam o grupo de itens específicos.

Segundo Reise et al. (2007), diferentemente dos modelos unidimensionais, existem poucos pacotes de softwares para realizar análise exploratória Bi-fatorial na TRI (o R é uma alternativa). No entanto, para itens dicotômicos é possível utilizar para estimação dos parâmetros os softwares NOHARM (FRASER, MCDONALD, 1988) e para casos de análise confirmatória e itens dicotômicos também, o software TESTFACT (WOOD et al., 2003).

## 2.4 SÍNTESE DO CAPÍTULO

O capítulo 2 tratou dos conceitos básicos que servem de estrutura teórica para o presente trabalho. O primeiro deles foi a definição de qualidade em websites comerciais a qual, apesar das inúmeras definições encontradas na literatura, pode ser generalizada segundo a classificação de Cristobal et al., (2007) e Loiacono (2002). Tal classificação conceitua qualidade em websites como um conjunto de características técnicas e não técnicas de um sistema web. Esta classificação converge diretamente aos objetivos do trabalho juntamente com as teorias de mediads discutidas a partir da seção 2.2 com os conceitos de criação de escalas e posteriormente,

na seção 2.3 com um aprofundamento sobre a ferramenta de medida estatística teoria da resposta ao item.

Um ponto importante destacado no levantamento bibliográfico sobre qualidade em websites é a característica multidimensional deste conceito discutido das 41 principais abordagens estudadas e classificadas. Dentre estas, encontrou-se abordagens práticas com no mínimo 3 e no máximo 15 dimensões. A maioria delas trabalhando com testes e percepção de usuários e utilizando como técnica de análise a análise fatorial exploratória e/ou confirmatória. As dimensões mais encontradas nestes estudos foram apresentação da informação, segurança, usabilidade e interatividade.

A seção 2.3 discute com mais detalhes o conceito de dimensionalidade, destacando a unidimensionalidade essencial e restrita e a multidimensionalidade entre itens e dentro de cada item.

Como alternativa para a criação de escalas e análise dimensional foi discutida a teoria da resposta ao item que é um conjunto de modelos probabilísticos que relaciona um traço latente de um respondente com a probabilidade de este responder a um item dentro de uma determinada categoria. Por se tratar de um construto multidimensional o foco desta seção foi o conjunto de modelos multidimensionais da TRI dentre eles o modelo compensatório de dois parâmetros e o bi-fatorial.

Este levantamento teórico serviu de arcabólso teórico para desenvolvimento da metodologia do presente trabalho para desenvolvimento de uma escala multidimensional para mensurar qualidade em webistes comerciais a qual é apresentada no próximo capítulo.

### 3. MÉTODO

A revisão da literatura apresentada no capítulo anterior contextualiza e apresenta as oportunidades de pesquisa no campo de mensuração de qualidade em websites com o objetivo de melhorar o entendimento das estruturas envolvidas no contexto de interação humano-computador. Neste capítulo é feito, primeiramente, uma breve caracterização da pesquisa, seguida de um detalhamento dos procedimentos metodológicos utilizados.

#### 3.1 CARACTERIZAÇÃO DA PESQUISA

Uma metodologia de pesquisa é designada para descrever os passos que são empregados para atingir os objetivos da pesquisa (TAYLOR; BOGDAN 1997).

Para isto, deve-se primeiro identificar e utilizar as ferramentas e técnicas adequadas, ou seja, a forma de abordagem do problema. Estas abordagens podem ser de natureza qualitativa ou quantitativa. A pesquisa quantitativa ocorre quando uma parte da investigação se baseia em análise estatística para tirar conclusões ou para testar hipóteses (ROMEY, 2007). As técnicas quantitativas normalmente provam ou refutam uma teoria específica que foi testada (BABBIE, 2005). Em outras palavras, em estudos quantitativos pesquisadores devem ter uma boa noção do que eles estão procurando, antes de iniciar seu estudo (NEILL, 2007).

A pesquisa qualitativa, por outro lado, baseia as suas conclusões sobre os debates, o pensamento e o conhecimento, a fim de ajudar a melhorar a compreensão de uma área de pesquisa (HART, 1999). Os métodos qualitativos são exploratórios ou de natureza descritiva e normalmente não são utilizados quando é necessário testar a teoria (BABBIE, 2005). Em outras palavras, a pesquisa qualitativa consiste em descobrir o que está acontecendo em uma determinada área de investigação. A diferença entre as abordagens quantitativa e qualitativa reside na natureza dos dados coletados e da forma em que estes são analisados. Atualmente é amplamente aceito que as duas abordagens se complementam (HORAN, 2010).

A presente pesquisa utiliza como base estes dois tipos de abordagem. A qualitativa, para gerar conhecimento sobre o contexto de

qualidade na web, utilizando revisão sistemática da literatura e a quantitativa, por meio da coleta de dados e do tratamento estatístico destes.

Além do tipo de abordagem do problema, segundo Silva e Menezes (2002) uma pesquisa pode ser classificada quanto à natureza, quanto aos objetivos e quanto aos procedimentos técnicos. Quanto à natureza, uma pesquisa pode ser classificada como básica ou aplicada. Segundo Hair et al. (2005), a pesquisa básica é motivada por um desejo de entender melhor algum fenômeno geral enquanto a pesquisa aplicada é motivado por uma tentativa de resolver um determinado problema específico. Desta forma, a presente pesquisa possui uma classificação quanto à natureza como aplicada, pois tem por objetivo compreender e modelar a qualidade em websites comerciais de uma forma.

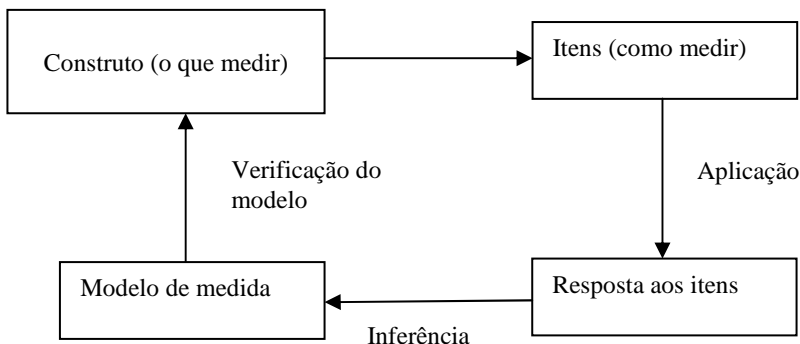
Quanto aos objetivos, Gil (1999) classifica a pesquisa como exploratória, descritiva ou explicativa. Esta pesquisa caracteriza-se como exploratória, pois visa proporcionar uma visão geral da qualidade em websites comerciais.

Quanto aos procedimentos técnicos, a pesquisa pode ser bibliográfica, documental, experimental, levantamento, estudo de caso, pesquisa ex-post-facto, pesquisa ação ou pesquisa participante (SILVA; MENEZES, 2000; GIL, 1999). A presente pesquisa utiliza os procedimentos técnicos bibliográficos, dos quais são retiradas as bases conceituais para desenvolvimento da modelagem, juntamente com o levantamento dos dados referentes à qualidade dos websites.

### 3.2 ELABORAÇÃO DOS ITENS

Como mencionado na seção anterior, a metodologia utilizada neste trabalho trata, primeiramente, de uma pesquisa exploratória com o objetivo de organizar o conhecimento na área de avaliação de qualidade em websites comerciais e, posteriormente, assume um caráter teórico aplicado, elaborando uma sistemática para mensurar qualidade em websites. A figura 8 esquematiza a sistemática utilizada.

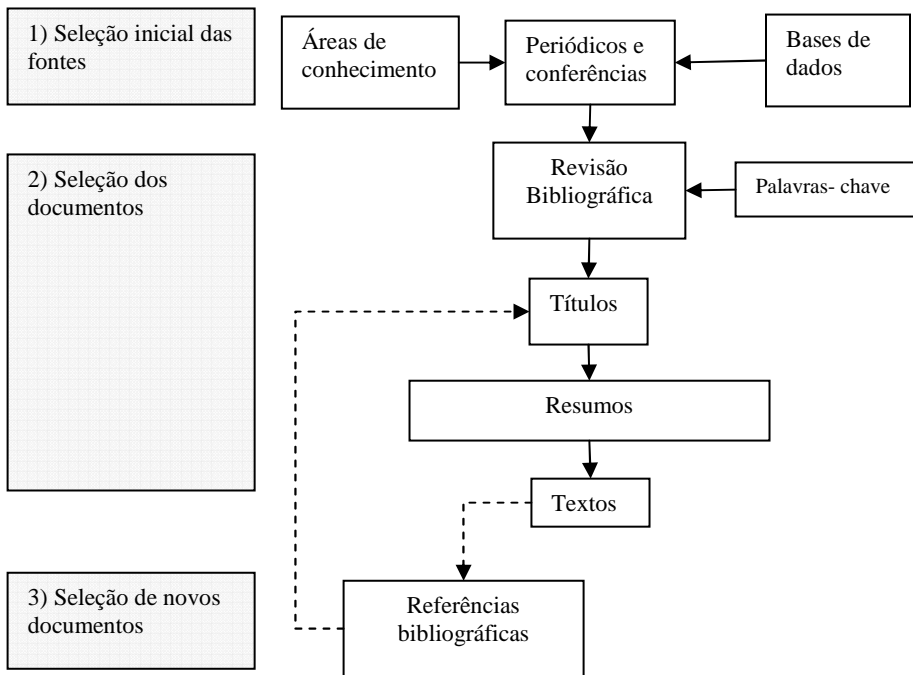
Figura 8 - Sistemática metodológica. Adaptada de Wilson (2005).



Inicialmente, definiu-se o que medir, ou seja, o construto: qualidade em websites comerciais. Para isso, foi realizada uma pesquisa exploratória para gerar conhecimento sobre este construto, esta pesquisa configurou-se inicialmente por um levantamento bibliográfico sistematizado como mostra a Figura 9.

No primeiro estágio, selecionaram-se as fontes iniciais, periódicos e anais de conferências. Nesta etapa, realizou-se o levantamento no portal de periódicos da CAPES (<http://www.periodicos.capes.gov.br/>) das publicações relacionadas às áreas de Ciências Sociais Aplicadas, Computação, Engenharias, Ciências da Informação e Multidisciplinares. Foram pesquisadas 18 bases de dados: Blackwell, Cambridge University Press, Emerald, Gale, IEEE/IET, INFORMS, Oxford University Press, Sage, Science Direct, Springer Verlag, Wilson, World Scientific, COMPENDEX Ei Engineering Index, CSA Cambridge Scientific Abstracts, EconLit (American Economic Association); Humanities Full Text (WilsonWeb), Scopus, Social Sciences Full Text, além de anais de congressos. No segundo estágio, para a seleção das referências, focou-se na busca por artigos entre os anos de 2000 e 2012 que continham as seguintes palavras-chave no título e/ou no resumo: “e-commerce” ou “web site” ou “website” ou “online shopping” ou “internet shopping” e “quality” ou “usability”. A partir daí, foram analisados os títulos e os resumos dos artigos encontrados, selecionando-se os que apresentarem relevância para os objetivos do trabalho. O terceiro estágio - seleção de novos documentos - é utilizado com o intuito de identificar, nas referências bibliográficas dos artigos encontrados no estágio 2, as mais citadas, com o objetivo de inseri-las na pesquisa, tal como mostra a retroalimentação na Figura 9.

Figura 9 - Método para o levantamento bibliográfico utilizado. Adaptado de Villas et al. (2008).



Ao todo, foram encontrados 191 artigos correlacionados com o tema mensuração de qualidade em websites. Após esta etapa, foi feita uma análise dos conceitos mais citados nos artigos levantados e suas associações, para o desenvolvimento do conjunto de itens a ser aplicado para medir qualidade em websites.

Segundo Wilson (2005), tipicamente, o item é mais uma das muitas maneiras que podem ser utilizadas para medir algum traço latente, como por exemplo, a qualidade em websites. Sob este ponto de vista, a criação de um instrumento é o resultado de uma série de decisões que o pesquisador necessita fazer considerando como será feita a representação do traço latente, qual metodologia de coleta de dados disponível, quanto tempo, comportamento da população, entre outros. Desta forma, como a presente pesquisa pretende representar o traço latente qualidade em websites por meio de uma escala multidimensional utilizando a teoria da resposta ao item, optou-se pela utilização de um conjunto de itens. Como mencionado no capítulo anterior, a construção destes itens parte de um levantamento rigoroso da literatura de modo que se pretende construir os itens com base



em estudos já realizados, embora existam, como discutido nos capítulos 1 e 2, divergências entre alguns estudos. O maior objetivo com a utilização desta metodologia para criação dos itens é tentar consolidar os diversos estudos, muitas vezes pontuais, presentes na literatura específica e desenvolver um instrumento abrangente e consistente.

Mais especificamente, a construção dos itens se deu por meio da associação dos conceitos resultantes da análise dos 191 artigos levantados no levantamento bibliográfico sistematizado, citado anteriormente. Por exemplo, alguns dos conceitos mais recorrentes foram “conteúdo da informação” associado com “credibilidade”, “exatidão”, “exaustividade”, “utilidade” o que é reforçado por Kim et al., (2005). Estas associações de conceitos deu suporte para a construção dos seguintes itens:

- existem informações básicas do produto? (conteúdo da informação + utilidade + credibilidade).
- o preço do produto consta ao lado da imagem ou link do produto? (conteúdo da informação + credibilidade + utilidade).
- os resultados de busca permitem classificação por outros critérios além de custos? (conteúdo da informação + utilidade + exatidão).

Alguns itens foram extraídos integralmente de estudos anteriores como, por exemplo, Tezza et al. (2011) e W3C (2008).

O planejamento do levantamento de dados segue a proposta de Stepchenkova et al (2010) e Tezza et al. (2011), no qual os itens são construídos para avaliar o sistema e não com o objetivo de captar percepção de clientes ou usuários, tendo portanto, a característica objetiva de um *checklist*. Com isso, trabalhou-se a elaboração de itens objetivos, capazes de verificar a existência ou não de determinada característica associada a qualidade do website comercial. Sendo assim, todos os itens foram formulados de forma objetiva e com respostas dicotômicas, ou seja, apresenta ou não a característica (0 ou 1), tendo em vista a aplicação destes e o tratamento estatístico posterior.

Foram elaborados 75 itens. Após a elaboração dos itens os mesmos foram submetidos à avaliação de 3 juízes, com o objetivo de verificar a adequação da representação comportamental dos itens ao traços latentes – qualidade em websites comerciais. Segundo Pasquali (1998), os juízes devem ser peritos na área do construto, pois sua tarefa consiste em ajuizar se os itens estão se referindo ou não ao traço em questão. Um dos juízes tem formação em design e trabalha com pesquisa em webdesign, outro é um profissional da área de computação e outro professor universitário da área de tecnologia da informação. Após a análise dos juízes, alguns ajustes foram feitos na formulação dos itens. O quadro 2 demonstra os 75 itens

finais assim como as fontes bibliográficas diretas, utilizadas para elaboração de cada item.

Quadro 2 - Descrição dos itens formulados e referidas referências diretas.

Item	Descrição do item	Fontes
01	Ao abrir a homepage há abertura de janelas pop-up?	Storey et al. (2002); Petre et al. (2006); Tezza et al. (2011)
02	A homepage apresenta as principais áreas de conteúdo do site (navegação)?	Nielsen (2000)
03	Homepage deixa claro o que o site faz (demonstra os principais produtos e/ou uma breve descrição de seu objetivo e/ou benefícios que oferece), sem precisar usar a rolagem?	Aparicio et al. (2006)
04	A homepage apresenta um resumo das promoções mais importantes?	Nielsen (2000)
05	O site possui imagens em movimento que podem distrair o usuário?	Colesca (2007)
06	Os links promocionais levam diretamente para a oferta anunciada?	Nielsen (2000)
07	Os menus estão organizados em ordem alfabética?	Olsina et al. (2001)
08	O site possui menu em cascata?	Nielsen; Loranger (2006)
09	Para navegação, existe uma trilha à esquerda ou links na parte superior?	Nielsen; Tahir (2002)
10	Existe agrupamento de subcategorias?	Schmidt et al. (2009)
11	O título da janela (browser) lista o nome do site em primeiro lugar?	Nielsen; Tahir (2002)
12	Existe informações de contato telefônico ou endereço?	Nielsen (2000); Nielsen; Tahir (2002); Aparício et al. (2006)
13	As páginas apresentam consistência visual, ou seja, mantém sempre a mesma aparência visual?	Aladwani; Palvia (2002); Ferreira; Nunes (2008)
14	O site possui a opção de acesso em outras línguas?	Colesca (2007)
15	As palavras clicáveis (coloridas ou sublinhadas) apresentam uma forma diferenciada quando são selecionadas?	Nielsen; Loranger (2006)
16	Os rótulos de campos começam com uma letra maiúscula, e as letras restantes são minúsculas?	Aldwani; Palvia (2002)
17	Os títulos estão alinhados à esquerda?	Wroblewski (2008); Inkpen et al. (2006)
18	Os parágrafos de texto são separados?	Zerfass; Hartmann,

Item	Descrição do item	Fontes
		(2005)
19	Palavras aparentemente clicáveis são de fato clicáveis?	Nielsen (2000)
20	Os títulos de telas, janelas e caixas de diálogo estão no alto, centrados ou justificados à esquerda?	Inkpen et al. (2006)
21	Todas as páginas possuem um campo de busca?	Nielsen (2000)
22	Quando há rolagem, não existem elementos de design (na tela inicial) que pareçam com marcadores de final de página?	Nielsen; Loranger (2006)
23	O logotipo da empresa está no canto superior esquerdo em todas as páginas do site?	Tan; Tung (2003); Roth et al. (2010)
24	Existe um link de um único clique que conduz a homepage?	van der Merwe; Bekker (2003)
25	O site permite navegação em suas páginas em apenas uma janela, ou seja, não há abertura de novas janelas em meio a navegação?	Nielsen; Loranger (2006); Tezza et al. (2011)
26	Há cores diferentes para links já visitados?	Nielsen; Tahir (2002)
27	Existe uma lista de perguntas frequentes - FAQs?	Colesca (2007); Hernández et al. (2007); Tezza et al. (2011)
28	Ao digitar uma pesquisa no campo de busca o buscador vai dando sugestões?	Long; McMellon (2004)
29	O sistema de busca é flexível quanto aos termos utilizados pelo usuário, ou seja, caso o usuário insira algum termo com grafia errada, o sistema de busca sugere correção?	Agner (2008)
30	Os resultados de busca permitem classificação por outros critérios além de custo?	Nielsen (2000); Tezza et al. (2011)
31	Listas longas apresentam indicadores de continuação, de quantidade de itens e de páginas?	Nielsen; Loranger (2006); Tezza et al. (2011)
32	Ícones de continuação de página estão visíveis?	Agner (2008)
33	O preço de um produto consta ao lado da imagem ou do link do produto?	Hung; McQueen (2004)
34	É possível ampliar as fotos dos produtos para visualizar detalhes?	Novikova (2009); Nielsen; Loranger (2006); Tezza et al. (2011)
35	Em produtos em que existam mais de uma perspectiva, é possível visualizar todas as perspectivas?	Nielsen; Tahir (2002); Tezza et al. (2011)
36	Os grupos de botões de comando estão dispostos em coluna e à direita, ou em linha e abaixo dos objetos aos quais estão associados?	Inkpen et al. (2006)

Item	Descrição do item	Fontes
37	Os produtos possuem informações suficientes (tamanho, características básicas)?	Signore (2005)
38	Existe a possibilidade de inserir feedback sobre os produtos por parte do consumidor?	Agner (2012)
39	Existem imagens de fundo nos textos?	Nielsen; Tahir (2002)
40	O site apresenta produtos relacionados no final da página?	Nielsen (2000)
41	Existe a opção de compartilhar a página nas redes sociais?	Agner (2012)
42	Existe a opção de ajuda on-line?	Hernández et al, (2007)
43	O site possui multimídia para apresentação de produtos?	Aladwani; Palvia (2002); Schmidt et al (2009)
44	A empresa oferece algum serviço grátis, como por exemplo, frete grátis?	Schmidt et al (2009)
45	Existe indicação de ambiente seguro no momento da efetivação da compra?	Nielsen; Tahir (2002)
46	No preenchimento de formulários, o usuário consegue visualizar os próximos passos na interface?	Nielsen (2000)
47	O site possui outras formas de pagamento além do cartão de crédito?	Yoo; Donthu (2001); Hung e McQueen (2004)
48	É possível saber os custos totais antes de fazer cadastro (inclusive frete)?	Nielsen; Loranger (2006)
49	Os campos destinados a formulários de preenchimento pelo usuário estão agrupados linearmente, evitando espaços desnecessários?	Inkpen et al. (2006) Tezza et al. (2011)
50	No preenchimento de um formulário, as opções que não são válidas ou não estão disponíveis estão visivelmente desativadas, prevendo assim erros?	Nielsen (2000)
51	No preenchimento de um formulário, é informado a forma de preenchimento?	Cybis (2007); Nielsen; Loranger (2006); Tezza et al. (2011)
52	Os dados obrigatórios são diferenciados dos dados opcionais de forma visualmente clara?	Cybis (2007); Nielsen; Loranger (2006)
53	É possível efetivar uma compra sem realizar um cadastro (que inclua login e senha)?	Nielsen; Loranger (2006); Tezza et al. (2011)
54	O sistema emite sinais sonoros quando ocorrem problemas na entrada de dados?	Nielsen; Loranger (2006); Tezza et al. (2011)

Item	Descrição do item	Fontes
55	As mensagens de erro estão isentas de abreviaturas e/ ou códigos gerados pelo sistema operacional?	Cybis (2007)
56	Qualquer ação do usuário pode ser revertida através da opção DESFAZER ou VOLTAR?	Nielsen; Loranger (2006); Tezza et al. (2011)
57	Todo o conteúdo não textual que é apresentado ao usuário tem uma alternativa em texto que serve um propósito equivalente?	W3C (2008)
58	As informações, a estrutura e as relações transmitidas através de apresentação podem ser determinadas de forma programática ou estão disponíveis no texto?	W3C (2008)
59	A cor não é utilizada como o único meio visual de transmitir informações, indicar uma ação, pedir uma resposta ou distinguir um elemento visual?	W3C (2008)
60	A apresentação visual de texto e imagens de texto tem uma relação de contraste de, no mínimo, 4,5:1?	W3C (2008)
61	Exceto para legendas e imagens de texto, o texto pode ser redimensionado sem tecnologia assistiva até 200 por cento sem perder conteúdo ou funcionalidade?	W3C (2008)
62	A apresentação visual do texto e imagens de texto tem uma relação de contraste de, no mínimo, 7:1?	W3C (2008)
63	Toda a funcionalidade do conteúdo é operável através de uma interface de teclado sem a necessidade de qualquer espaço de tempo entre cada digitação individual, exceto quando a função subjacente requer entrada de dados que dependa da cadeia de movimento do usuário e não apenas dos pontos finais?	W3C (2008)
64	Para cada limite de tempo definido pelo conteúdo, no mínimo, uma das seguintes afirmações é verdadeira: - Desligar: O usuário pode desligar o limite de tempo antes de o atingir; ou -Ajustar: O usuário pode ajustar o limite de tempo antes de o atingir, acima de um grande intervalo que dure, no mínimo, dez vezes mais do que a predefinição; ou -Prolongar: O usuário é avisado antes de o tempo expirar e tem, no mínimo, 20 segundos para prolongar o limite de tempo com uma simples ação (por exemplo, "pressionar a barra de espaços"), e o usuário pode prolongar o limite de tempo, no mínimo, dez vezes; ou -Excepção em Tempo Real: O limite de tempo é uma parte necessária de um evento em tempo real (por exemplo, um leilão), e não é possível nenhuma alternativa ao limite de tempo; ou -Excepção Essencial: O limite de tempo é essencial e prolongá-lo iria invalidar a atividade; ou -Excepção de 20 Horas: O limite de tempo é superior a	W3C (2008)

Item	Descrição do item	Fontes
	20 horas.	
65	<p>Para informações em movimento, em modo intermitente, em deslocamento ou em atualização automática, todas as seguintes afirmações são verdadeiras:</p> <ul style="list-style-type: none"> <li>- Em movimento, em modo intermitente, em deslocamento: Para quaisquer informações em movimento, em modo intermitente ou em deslocamento, que (1) sejam iniciadas automaticamente, (2) durem mais de cinco segundos, e (3) sejam apresentadas em paralelo com outro conteúdo, existe um mecanismo para o usuário colocar em pausa, parar, ou ocultar as mesmas, a menos que o movimento, o modo intermitente ou o deslocamento façam parte de uma atividade, na qual sejam essenciais; e</li> <li>- Em atualização automática: Para quaisquer informações em atualização automática, que (1) sejam iniciadas automaticamente e (2) sejam apresentadas em paralelo com outro conteúdo, exista um mecanismo para o usuário colocar em pausa, parar ou ocultar as mesmas, ou controlar a frequência da atualização, a menos que a atualização automática faça parte de uma atividade, na qual é essencial.</li> </ul>	W3C (2008)
66	Está disponível um mecanismo para ignorar blocos de conteúdo que são repetidos em várias páginas Web?	W3C (2008)
67	As páginas Web têm títulos que descrevem o tópico ou a finalidade?	W3C (2008)
68	A finalidade de cada link pode ser determinada a partir apenas do texto do link ou a partir do texto do link juntamente com o respectivo contexto do link determinado de forma programática, exceto quando a finalidade do link for ambígua para os usuários em geral?	W3C (2008)
69	Os cabeçalhos e as etiquetas descrevem o tópico ou a finalidade?	W3C (2008)
70	A linguagem humana pré-definida de cada página Web pode ser determinada de forma programática?	W3C (2008)
71	Alterar a definição de um componente de interface de usuário não provoca, automaticamente, uma alteração de contexto, a menos que o usuário tenha sido avisado sobre essa situação antes de utilizar o componente?	W3C (2008)
72	Os componentes que têm a mesma funcionalidade num conjunto de páginas Web são identificados de forma consistente?	W3C (2008)
73	Se um erro de entrada for automaticamente detectado, o item que apresenta erro é identificado e o erro é descrito ao usuário por texto?	W3C (2008)
74	Etiquetas ou instruções são fornecidas quando o conteúdo	W3C (2008)

Item	Descrição do item	Fontes
	exigir a entrada de dados por parte do usuário?	
75	No conteúdo implementado utilizando linguagens de marcação, os elementos dispõem de marcas de início e de fim completas, os elementos estão encaixados de acordo com as respectivas especificações, os elementos não contêm atributos duplicados, e todos os IDs são exclusivos, exceto quando as especificações permitem estas características?	W3C (2008)

### 3.3 COLETA DOS DADOS

O conjunto de 75 itens gerados e ajustados com base nas sugestões dos juízes foi aplicado em uma amostra de 441 sites de e-commerce de origem brasileira.

Para definição da amostra, foi utilizada amostragem intencional, procurando extrair da população de websites comerciais de origem brasileira uma amostra com websites de baixa, média e alta qualidade. Para isso, além de sites com os mais diversos gêneros de produtos comercializados, observou-se a variedade de estilos de design, usabilidade, estética e layout contemplando desde algo relativamente primário até excessivamente elaborado, o que não necessariamente implica em alta ou baixa qualidade, apenas garante a diversidade, pré-requisito da teoria da resposta ao item. Não existe um consenso sobre o tamanho ideal de amostras para utilização da teoria da resposta ao item (DOWNING, 2003, WONGTADA; RICE, 2008). Segundo Spencer (2004) e Watson et al. (2006), modelos multidimensionais da teoria da resposta ao item requerem uma amostra de pelo menos 400 respondentes. Wongtada e Rice (2008) argumentam que grandes amostras aumentam artificialmente os valores das estatísticas, mas pequenas amostras reduzem o poder estatístico do teste. Che et al. (2006) aplicam um modelo bi-fatorial em uma amostra de 403 indivíduos, Wolfe et al. (2006) utiliza uma amostra de 567 indivíduos para aplicação do modelo multidimensional MRCMLM e, Wongtada e Rice (2008) utilizam 473 indivíduos para aplicar o modelo multidimensional para medir percepção de inovação organizacional. Sendo assim, optou-se por trabalhar com uma amostra de 441 websites.

A coleta desta amostra foi feita com o auxílio de buscadores como Google.com, Bing.com e Yahoo.com no período de janeiro e fevereiro de 2012.

O Quadro 3 demonstra a composição da amostra segundo o tipo de produto comercializado pelo site, as URLs dos websites podem ser vizualizadas no Apêndice 1.

Quadro 3 - Composição da amostra segundo o tipo de produto comercializado pelo website.

<b>Produto</b>	<b>Sites (quantidade)</b>	<b>Sites (%)</b>
Acessórios automotivos	18	4,08
Adesivos	13	2,95
Alimentos e Bebidas	34	7,71
Artigos esportivos	27	6,12
Bolsas e acessórios	6	1,36
Brinquedos	11	2,49
Calçados	8	1,81
Cama mesa e banho	11	2,49
Cosméticos	11	2,49
Drogaria	12	2,72
DVD/CD	8	1,81
Eletrônicos	21	4,76
Floricultura	14	3,17
Informática	8	1,81
Instrumentos musicais	19	4,31
Lentes	10	2,27
Livraria	25	5,67
Loja de departamentos	36	8,16
Móveis	9	2,04
Outros	13	2,95
Perfumes	22	4,99
Pet Shop	6	1,36
Produtos específicos/personalizados	24	5,44
Produtos exotéricos/religiosos	14	3,17
Sex Shop	20	4,54
Tabacaria	11	2,49
Vestuário	30	6,80
Total	441	100,00

Dos 75 itens gerados, 56 deles tiveram suas respostas obtidas por meio de coleta manual, na qual o aplicador acessa o site em análise e



responderá, após rápida navegação no website, se o mesmo possui ou não a características em questão. Este método manual de coleta de dados em websites baseia-se em Pinterits et al. (2006), Al-Khalifa (2010), Stepchenkova et al. (2010) e Tezza et al. (2011). A coleta manual foi realizada pelo pesquisador e uma bolsista que recebeu um treinamento de cerca de uma hora para entender os objetivos da pesquisa e a metodologia da coleta de dados. Esta coleta manual levou dois meses – março a maio de 2012.

O pesquisador e a bolsista avaliaram cerca de 200 websites e mais 50 websites comuns para garantir que as avaliações produzidas por ambos os avaliadores fossem consistentes e precisas. Uma avaliação pré-teste foi realizada, no qual os avaliadores avaliaram 5 websites escolhidos aleatoriamente. Os resultados mostraram que os avaliadores tinham necessidade de ajustes avaliativos em cerca de 10 itens da pesquisa, para poder sincronizar as avaliações. Diretrizes de avaliação foram feitas e perguntas foram reescritas e amplamente discutidas para garantir a compreensão comum. Os 50 websites avaliados por ambos os avaliadores foram sorteados aleatoriamente pelo orientador deste trabalho e divididos em dois grupos um de 20, e outro de 30 e suas avaliações foram espaçadas ao longo de todo o processo de coleta de dados. Esta estruturação metodológica foi baseada em Krippendorff (2004) e Stepchenkova et al (2010). Os avaliadores não conheciam *a priori* quais os websites eram comuns, para evitar alterações no processo avaliativo. Após a coleta do primeiro lote de websites comuns (20 websites) verificou-se a correlação entre as respostas, para isso utilizou-se o coeficiente de contingências, por se tratar de respostas qualitativas nominais. Nesta etapa encontrou-se um coeficiente de contingência de 0,77 significando que em média, somente 2,8 itens foram avaliados de forma diferente. Resultado similar foi verificado ao final da coleta, obtendo-se um coeficiente de 0,79, significando uma discordância média de menos de 2 itens por site.

O ambiente para coleta de dados foi constituído por um computador com tela de 15 polegadas. O navegador utilizado foi o Internet Explorer versão 7, por ser, o navegador mais utilizado no mundo atualmente. A resolução da tela utilizada foi de 1024 por 768 pixels que, segundo Nielsen e Loranger (2006 p.225), “é o tamanho mais popular e a tendência aponta para 1024 por 768 como o menor tamanho de tela no futuro”.

Os 19 itens restantes foram analisados de forma semiautomática utilizando a ferramenta livre AChecker (<http://achecker.ca/>). Esta ferramenta verifica as diretrizes de acessibilidade com o objetivo de estabelecer o nível de acessibilidade de websites ao detectar erros de acordo

com as diretrizes de Acessibilidade para Conteúdo Web (WCAG) 2.0. A WCAG 2.0, é um conjunto de recomendações desenvolvida pela W3C (Consórcio World Wide Web) estas diretrizes tem por objetivo verificar a acessibilidade do conteúdo de uma interface, considerando o usuário com alguma deficiência ou limitação e melhorar a relação dos usuários em geral.

A ferramenta AChecker funciona online. Primeiramente é inserido o endereço do website que se deseja avaliar, sendo possível, antes de solicitar a avaliação, configurar o tipo de análise, ou seja, quais recomendações utilizar. Além da WCAG 2.0, a ferramenta pode avaliar a WCAG 1.0, a BITV 1.0, Section 508 ou a Stanca Act. Foi utilizada a WCAG 2.0 em todos os seus níveis. A análise pode levar alguns minutos, dependendo do tamanho da página, ao final da análise a ferramenta gera um relatório, no qual traz as recomendações da WCAG 2.0 que não são atendidas e sugere alterações que podem melhorar a interface para que consiga atender cada uma das recomendações. Como os 19 itens avaliados de foram semiautomática forma extraídos das recomendações de acessibilidade da WCAG 2.0, por meio do relatório gerado em cada análise, foi verificado quais itens não eram atendidos pelo websites e consequentemente quais eram, representando, portanto, uma análise semiautomática feita pela ferramenta AChecker e pelo autor com respostas do tipo dicotômicas.

### 3.4 METODOLOGIA ESTATÍSTICA

Na etapa de análise dos dados, foram realizadas análises de dimensionalidade, realizando-se uma análise exploratória para identificação do número de dimensões e a adequação de cada item às dimensões. Foi verificado a adequação do modelo multidimensional compensatório da teoria da resposta ao item e por fim, com base na análise exploratória, foi realizada a verificação do modelo bifatorial.

Esta etapa seguiu os seguintes passos:

- avaliação do número de dimensões por meio de análise fatorial de informação restrita e de informação plena;
- avaliação da qualidade dos 75 itens, com base nas cargas da análise fatorial de informação plena;
- validação das dimensões sugeridas na primeira etapa por meio de análise estatística e de relação do conteúdo de cada item nas referidas dimensões;
- verificação da adequação do modelo de multidimensional compensatório da teoria da resposta ao item frente ao modelo unidimensional; e

- verificação da adequação do modelo bi-fatorial frente ao modelo multidimensional compensatório da teoria da resposta ao item e ao modelo unidimensional.

### 3.4.1 - Avaliação do número de dimensões

A avaliação no número de dimensões do conjunto inicial de dados, com 75 itens, foi feita por meio do método de análise fatorial de informação restrita e o método de análise fatorial de informação plena. No primeiro método, o número de dimensões foi verificado com base na matriz de correlação tetracórica e na análise paralela discutida na seção 2.3.2, para isso foi utilizado o pacote psych (REVELLE, 2012) implementado no software R (R CORE TEAM, 2012).

Como contraponto, a dimensionalidade do conjunto total foi verificada por meio do método de informação plena. Neste método, conforme discutido na seção 2.3.2, a natureza dicotômica dos itens limita o uso da análise fatorial exploratória mais frequentemente encontrada na literatura (JOHNSON, WICHERN, 2007) e implementada nos programas estatísticos tradicionais. Foi utilizada a abordagem descrita por Bock e Aitkin (1981) e Bock, Gibbons e Muraki (1988), na qual o tratamento de itens dicotômicos e estimação das cargas dos fatores é feito por meio da técnica denominada análise fatoriais de informação plena, implementada no software R (R CORE TEAM, 2012) no pacote mirt (CHALMERS, 2012) e no software flexMIRT<sup>TM</sup> (CAI, 2012). Neste método, o número de dimensões foi avaliado com base em dois critérios de informação, o critério de informação bayesiano (*Bayesian Information Criterion* – BIC) (SCHWARZ, 1978) e o critério de informação de Akaike (*Akaike Information Criterion* – AIC) (AKAIKE, 1973). A utilização deste método para determinação do número de dimensões é discutido por Bartolucci et al. (2012), Nylund et al. (2007) e Rost (1997). Basicamente estes critérios comparam o possível aumento de informação em um modelo com número de dimensão  $m$  frente a um modelo com número de dimensões  $m+1$ , isto é feito como base no valor de máxima verossimilhança do modelo e do número de parâmetros livres (graus de liberdade). Assume-se que o melhor modelo é o que apresenta menor valor nestes critérios (BARTOLUCCI et al., 2012; ROST, 1997).

### 3.4.2 - Avaliação da qualidade dos itens

Após a verificação do número de dimensões do conjunto de 75 itens, realizou-se a avaliação da qualidade dos itens. Teve como objetivo principal reduzir o número de itens do conjunto com base na carga fatorial de cada item. Assumiu-se que itens com carga fatorial muito baixa em todas as dimensões refletem em itens que não possui relação direta com o construto que esses estão medindo, no caso, qualidade em websites comerciais. Itens com carga inferior a 0,300 ou comunalidade inferior a 0,200 foram considerados de baixa qualidade e consequentemente foram retirados do conjunto, seguindo a indicação de Hair et al. (2009).

### 3.4.3 - Validação das dimensões

Após a redução do número de itens por meio da metodologia descrita no item anterior, foi feita uma nova verificação do número de dimensões para os itens restantes. Esta avaliação, num primeiro momento seguiu exatamente a mesma metodologia da subseção 3.4.1; num segundo momento foi avaliada a possível correlação entre as dimensões, com a finalidade de definir a escolha no tipo de rotação a ser utilizada para distribuição das cargas dos itens em cada dimensão. A partir desta definição, foram avaliadas as cargas de cada item nas dimensões, posicionando assim o item na dimensão em que este apresenta maior carga. Ao final desta distribuição, cada grupo de itens característicos de uma dada dimensão foi avaliado quanto ao seu conteúdo tendo como base a teoria sobre qualidade em websites.

### 3.4.4 - Verificação da adequação do modelo unidimensional e unidimensional múltiplo

O primeiro modelo, relacionado diretamente com a teoria da resposta ao item, testado foi o modelo unidimensional de dois parâmetros, representado pela equação (4). O objetivo principal da aplicação deste modelo em um construto, teoricamente, multidimensional foi verificar o comportamento individual de cada item e identificar algum padrão que pudesse identificar grupos de itens com comportamento unidimensional. Além desta modelagem puramente unidimensional aplicada a todos os itens, para efeito de comparação com o modelo multidimensional, foi aplicado um modelo unidimensional em  $k$  grupos de itens, denominado neste trabalho de modelo unidimensional múltiplo não correlacionado. A definição do número de grupos ( $k$ ) foi feita com base na definição do

número de dimensões validadas conforme metodologia indicada na seção anterior.

#### 3.4.5 - Verificação da adequação do modelo de multidimensional compensatório

Apesar de haver um paralelo direto entre análise fatorial e a Teoria da Resposta ao Item Multidimensional (TRIM), como discutido na seção 2.3.1, a TRIM desenvolve uma análise mais completa, considerando além das cargas fatoriais, tidas no modelo como parâmetros de discriminação dos itens nas dimensões, o posicionamento relativo dos itens em relação a sua dificuldade e, além disso, estima a habilidade do respondente na mesma métrica, que na presente abordagem é o grau de qualidade de um website comercial. Desta forma, a comparação entre a AF e a TRIM se deu apenas no nível de carregamento das cargas fatoriais e dos parâmetros de discriminação.

Para verificar a adequação do modelo multidimensional compensatório realizou-se primeiramente uma comparação estatística entre o modelo unidimensional da TRI, o modelo logístico de dois parâmetros e o modelo multidimensional compensatório. Esta primeira comparação foi feita comparando-se os dois critérios de informação descritos da subseção 3.4.1, o critério de informação bayesiano (BIC) (SCHWARZ, 1978) e o critério de informação de Akaike (AIC) (AKAIKE, 1973), juntamente com o teste de análise de variância (Anova) sugerido e implementado no R por Chalmers (2012). Ainda nesta comparação, realizou-se uma verificação teórica levando em consideração os conceitos de qualidade de website contida em cada item e a estimação dos parâmetros em cada modelo.

Uma segunda comparação foi realizada, assumindo-se um modelo unidimensional múltiplo não correlacionado multi-unidimensional, ou seja, após identificação dos itens nas  $k$  dimensões encontradas na subseção 3.4.3, definiram-se  $k$  modelos unidimensionais e avaliou-se a consistência teórica das estimações comparando os  $k$  modelos unidimensionais com o modelo multidimensional.

#### 3.4.6 - Verificação da adequação do modelo bi-fatorial

A diferença básica entre o modelo bi-fatorial e o modelo multidimensional compensatório da TRI, é a de que o primeiro apresenta um fator geral no qual todos os itens carregam e outras dimensões específicas. Mais detalhes sobre as diferenças entre estes modelos foram discutidos no capítulo 2.

O modelo bi-fatorial mais difundido na literatura é o modelo de natureza confirmatória. Casos específicos de análise exploratória Jenrich e Bentler (2011), vêm sendo desenvolvidos; entretanto, sua aplicação prática ainda é limitada. Desta forma, a abordagem adotada neste trabalho foi a confirmatória. Para isso, foram tomadas como base as definições do número de dimensões e dos agrupamentos de itens definidos na seção 3.4.3 assumindo-se como dimensão geral a qualidade de websites comerciais.

O ajuste do modelo bi-fatorial com relação ao modelo compensatório da TRIM foi avaliado com base nos critérios de informação descritos da subseção 3.4.1, o critério de informação bayesiano (BIC) (SCHWARZ, 1978) e o critério de informação de Akaike (AIC) (AKAIKE, 1973).

Além da avaliação do ajuste estatístico, foi verificado o ajuste teórico dos itens às duas estruturas, bi-fatorial e multidimensional compensatória, e também a avaliação dos escores (grau qualidade dos websites) e seus significados práticos em cada escala gerada.

### 3.5 SÍNTESE DO CAPÍTULO

O procedimento metodológico utilizado no presente trabalho, de uma forma geral, contou com a caracterização da pesquisa, elaboração dos itens, coleta de dados e metodologia estatística.

Quanto à caracterização, a presente pesquisa é predominantemente quantitativa contando, entretanto, com uma base exploratória qualitativa que teve como objetivo compreender a área de estudo sobre qualidade em websites comerciais e servir de base para a elaboração dos itens.

Foram elaborados e validados por juízes 75 itens, todos teoricamente ligados a qualidade de websites. Tais itens, apesar de serem itens de resposta objetiva e independentes da percepção dos usuários foram baseados em estudos anteriores que utilizaram testes com usuários e/ou pesquisa de satisfação.

A coleta de dados foi realizada em 441 websites comerciais de origem brasileira. Dos 75 itens, 56 foram coletados de forma manual e outros 19 foram coletados automaticamente por meio da ferramenta Achecker.

A metodologia de análise estatística realizou primeiramente avaliação do número de dimensões do conjunto de itens seguido de uma verificação da qualidade de cada item seguindo assim para uma validação das dimensões e a verificação da adequação dos modelos multi e

unidimensionais da teoria da resposta ao item assim como o bi-fatorial. Os resultados e as discussões destas verificações são apresentados no capítulo seguinte.

## 4. RESULTADOS E DISCUSSÕES

Neste capítulo são apresentados os resultados das análises descritas no capítulo 1 e no capítulo 3 juntamente com uma discussão destes resultados tendo como eixos de análise a teoria estatística e a teoria sobre qualidade de websites.

### 4.1 AVALIAÇÃO DO NÚMERO DE DIMENSÕES

Na primeira análise para avaliar o número de dimensões utilizaram-se os 75 itens descritos no Quadro 2 e aplicou-se o método da análise fatorial de informação restrita, baseada na correlação tetracórica e na análise paralela.

O primeiro passo para avaliar a dimensionalidade utilizando o método de informação restrita consiste na inspeção dos valores próprios da matriz de correlação tetracórica (COSTA, 2011). Na Tabela 1, são apresentados os primeiros dez autovalores da matriz de correlação tetracórica e na figura 10 a respectiva representação gráfica.

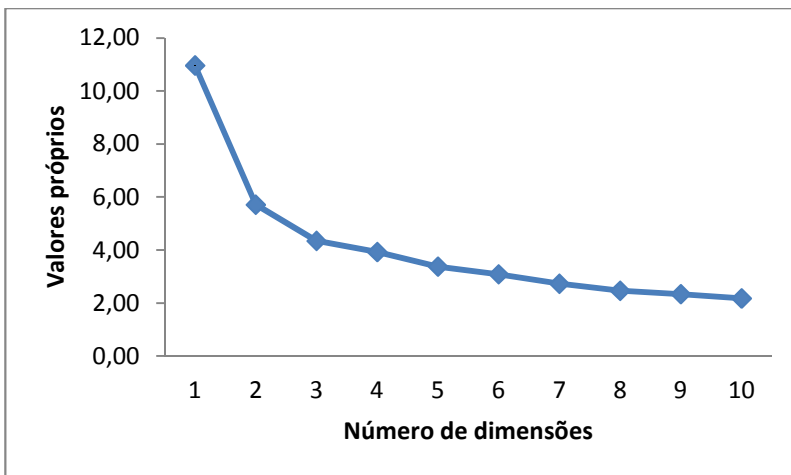
Tabela 1 - Valores próprios da matriz de correlação tetracórica.

Dimensão	1	2	3	4	5	6	7	8	9	10
<b>Autovalor</b>	11,0	5,7	4,4	3,9	3,4	3,1	2,7	2,5	2,3	2,2
<b>Proporção acumulada da variação explicada</b>	14,6	22,2	28,0	33,3	37,8	41,9	45,5	48,8	51,9	54,8

Verifica-se que o primeiro autovalor é 11,0, o que em um conjunto de 75 itens representa 14,6% da variação total explicada pelo primeiro fator ou primeira dimensão. Este resultado traz evidências de que o construto não deve ser assumido como unidimensional.



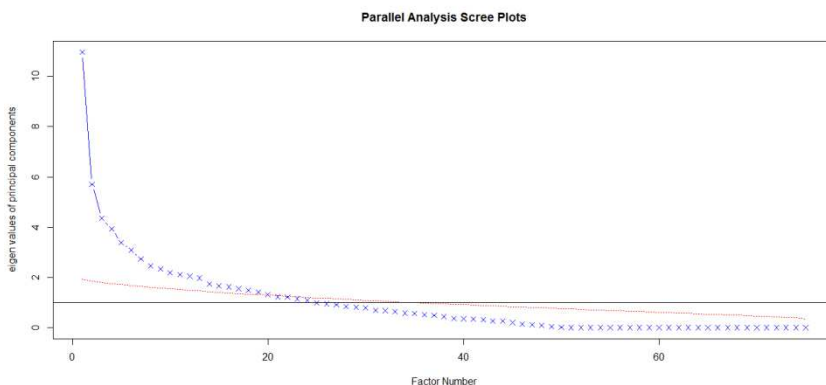
Figura 10 - Valores próprios da matriz de correlação tetracórica considerando 10 dimensões para o fator latente.



Quanto à definição do número exato de dimensões, existe na literatura algumas indicações de análise, como, por exemplo, o critério de que pelo menos 50% da variância total dos dados originais devem ser explicados por estes fatores (FERREIRA et al. 2010) representando neste caso 9 dimensões como pode ser visualizado na última linha da tabela 1. Outro possível critério é o de Kaiser (KAISER, 1958), o qual considera apenas os autovalores superiores a 1 (um), demonstrando que esses seriam os valores estatisticamente significativos. No entanto, segundo alguns autores, como, por exemplo, Fleck e Bourdel (1998), esta condição não é suficiente. Nem todos os autovalores superiores a um correspondem a componentes com significado evidente. Além disso, segundo Bryman e Cramer (1992), o critério de Kaiser tem sido recomendado para situações em que o número de variáveis é menor do que 30 e a média das comunalidades é maior do que 0,700; ou quando o número de indivíduos da amostra é superior a 250, e a comunalidade média é maior ou igual a 0,600. A consideração destes elementos sugeriu a inaplicabilidade desse critério ao conjunto dos 75 itens.

O método da análise paralela, conforme descrito na seção 2.3.2, indica a existência de 24 dimensões. Esta conclusão pode ser verificada no gráfico da Figura 11 onde a linha pontilhada refere-se ao conjunto de dados simulados e a linha cheia representa os dados reais. Observa-se que existem 24 autovalores acima da linha pontilhada.

Figura 11 - Resultado da análise paralela para os 75 itens.



Como uma forma de se complementar a análise feita pelo método de informação restrita, utilizou-se a análise fatorial com o método de informação plena. Segundo esta abordagem, a primeira dimensão explica apenas 12,4% da variação total, o que caracteriza, assim como a análise de informação restrita, um construto com mais de uma dimensão.

Os *scree plots* apresentados nas Figuras 10 e 11 sugerem que há pouca correlação entre os itens. Uma situação em que é difícil resumir as informações das respostas aos itens por poucos fatores subjacentes, o que é confirmado pela existência de mais de um fator. A correlação média, em módulo, entre os 75 itens foi de 0,15 e destas, pouco menos de 3% apresentaram correlação maior que 0,50.

Segundo Chalmers (2012), o número de dimensões que gera um melhor ajuste aos dados pode ser verificado por meio de uma comparação de modelos utilizando uma análise de variância genérica (ANOVA) implementada no software R. O resultado retorna o qui-quadrado ( $\chi^2$ ) baseado no teste de verossimilhança, bem como a diferença no AIC e BIC na comparação dos modelos. Foi feita a comparação de quatro modelos, o primeiro assumindo uma dimensão (Mod1), o segundo, duas (Mod2), o terceiro, três (Mod3) e o quarto, quatro (Mod4). A tabela 2 demonstra os resultados.

Tabela 2 - Comparação dos modelos de um, dois, três e quatro dimensões.

Modelos	Diferença qui-quadrado			AIC	BIC
	$\chi^2$	Graus de liberdade	p-valor		
Mod1 x Mod2	525,50	74	<0,01	377,50	74,91
Mod2x Mod3	283,32	73	<0,01	137,32	-161,18
Mod3 x Mod4	83,88	72	0,16	-60,12	-354,53

O teste estatístico assume como hipótese nula que não há diferença de informação entre os modelos, ou seja, assumir duas dimensões não agrega informação significativa ao modelo em comparação ao modelo de uma dimensão, e a hipótese alternativa assume que há ganho significativo na informação. Verifica-se, na Tabela 2, que a diferença entre o modelo 1 e o modelo 2 é estatisticamente significativa para  $\alpha = 0,05$ , (hipótese alternativa aceita) indicando que o modelo que assume duas dimensões possui mais informação que o modelo de uma dimensão, isso também pode ser visualizado nos critérios de informação AIC e BIC, nos quais verifica-se que existe diferenças positivas do modelo 1 em relação ao modelo 2. O mesmo ocorre na comparação do modelo 2 e do modelo 3, no entanto, neste caso, verifica-se no AIC que há diferença positiva entre os modelos 2 e 3 mas, no BIC o mesmo não ocorre. Estes dois critérios já foram comparados em vários estudos (DIAS, 2006; NYLUND et al., 2007, McLACHLAN; PEEL, 2000) e, segundo Bartolucci et al. (2012) o AIC eventualmente tende a superestimar o número de dimensões. Já o BIC, em alguns casos, pode subestimar.

Na comparação entre o modelo de três dimensões e o modelo de quatro dimensões não é apresentada diferença significativa, indicando que a utilização de um modelo de quatro dimensões pode não fornecer mais informação do que um modelo de três dimensões. Portanto, o teste de diferença de  $\chi^2$  indica que o conjunto de dados pode ser tratado como um modelo contendo três fatores. O teste de diferença de  $\chi^2$  para comparação de modelos possui algumas fragilidades inerentes ao equacionamento matemáticos que estão associados ao número de respondentes, ao número de parâmetros envolvidos, a existência de dados omitidos, entre outros. Por isso, seus resultados são tomados aqui como um indicador que necessita de respaldo teórico com base na literatura de qualidade de websites.

A mesma conclusão é obtida ao se comparar, conforme recomendação de Bartolucci et al. (2012), a quantidade de informação medida em cada um dos modelos individualmente utilizando os critérios BIC e AIC. Estas comparações podem ser visualizadas na tabela 3. Nota-se

que as diferenças entre os valores de AIC e BIC entre os modelos são as mesmas apresentadas nas duas últimas colunas da Tabela 2.

Tabela 3 - Índice de AIC e BIC para os modelos de um, dois, três e quatro dimensões.

Modelo	-2 Log da máxima verossimilhança	AIC	BIC
Mod1	-12555,47	25610,94	26633,20
Mod2	-12292,72	25233,44	<b>26558,29</b>
Mod3	-12151,06	<b>25096,12</b>	26719,47
Mod4	-12109,12	25156,24	27074,00

Por meio da análise da Tabela 3, verifica-se que, pelo critério AIC, há três dimensões no construto, já o critério BIC indica a existência de 2 dimensões, neste caso, com base no teste de significância mostrado na tabela 2, verifica-se que o critério de AIC para esta análise é o mais adequado pois é reforçado pelo teste  $\chi^2$  apresentado na Tabela 2.

Sendo assim, as análises estatísticas realizadas não são conclusivas quanto ao número exato de dimensões, indicando, entretanto, que possivelmente o conjunto de itens possui de 3 a 24 dimensões. Alguns autores, como McDonald (1981) e Spector et al. (1997) argumentam que a análise fatorial clássica tende a superestimar o número de dimensões, criando fatores espúrios quando não há uniformidade na dificuldade dos itens. Desta forma não é muito conveniente considerar um número muito alto de dimensões tendo em vista também no número limitado da amostra. Sendo assim, neste caso, um critério empírico, recomendado por diferentes autores, como Ford et al (1986), Fleck e Bourdel (1998), Podsakoff, Ahearne e MacKenzie (1997), pode ser utilizado. Este método propõem conservar as dimensões das quais sabe-se claramente dar um significado. Antes de iniciar uma análise empírica das dimensões, sugere-se avaliar a qualidade dos itens com o objetivo de eliminar itens que não estejam medindo os construtos de forma coerente com o grande conjunto de itens do instrumento e com isso estejam dando origens a possíveis dimensões espúrias (VELDKAMP, van der LINDER, 2002; FINKELMAN et al., 2010), os resultados deste procedimento são descritos e discutidos na próxima seção.

#### 4.2 AVALIAÇÃO DA QUALIDADE DO CONJUNTO DE ITENS INICIAL

Na seção anterior identificou-se que a definição do número de dimensões do conjunto inicial, contendo 75 itens, foi inconclusível.

Indicando, entretanto, a presença de no mínimo três dimensões e possivelmente mais de 10 dimensões. Uma das alternativas para tratar esta indefinição pode ser a análise empírica dos itens e das dimensões. No entanto, antes de iniciar uma análise empírica é importante lembrar que os 75 itens foram gerados com base na literatura, validado por especialistas, mas, sem necessariamente passar por uma análise quantitativa da qualidade de cada item após a coleta dos dados. Esta análise pode ser feita com base na verificação das cargas fatoriais de cada item, o que reflete a relação do item com os fatores subjacentes (traços latentes) presentes no conjunto de dados.

A análise das cargas e dos itens pode ser feita com base em alguma técnica de rotação para que as relações entre as dimensões e os itens possam ser evidenciadas e melhor interpretadas. Estas rotações podem ser ortogonais, quando os fatores possuem baixa correlação, ou oblíquas, quando os fatores são correlacionados. Para decidir qual das técnicas utilizar, verificou-se a correlação dos fatores assumindo 3 dimensões, Tabela 4.

Tabela 4 - Correlação dos fatores assumindo três dimensões aos 75 itens.

	F_1	F_2	F_3
F_1	1,000	0,232	0,020
F_2	0,232	1,000	0,069
F_3	0,020	0,069	1,000

Observa-se, na Tabela 4, que os fatores possuem uma baixa correlação. Sendo assim, optou-se por utilizar uma rotação ortogonal. A Tabela 5 mostra a análise fatorial utilizando o método da informação plena, e fazendo a rotação ortogonal varimax.

Tabela 5 - Análise fatorial de informação plena, com rotação ortogonal varimax.

Item	Dimensões			Item	Dimensões		
	1	2	3		1	2	3
01	0,015	-0,328	0,167	39	-0,157	-0,255	0,210
02	0,236	0,634	-0,152	40	0,539	-0,109	0,070
03	0,534	-0,049	-0,124	41	0,350	0,280	-0,059
04	0,277	0,118	0,098	42	0,292	-0,108	-0,002
05	-0,112	-0,501	0,356	43	0,305	-0,374	0,101
06	0,534	-0,362	-0,204	44	0,437	0,031	0,106

Item	Dimensões			Item	Dimensões		
	1	2	3		1	2	3
07	-0,159	-0,231	0,082	45	0,645	-0,257	0,070
08	-0,485	-0,448	-0,178	46	0,519	0,551	0,219
09	0,313	-0,050	-0,050	47	0,237	-0,063	-0,395
10	0,454	-0,052	0,277	48	0,792	-0,145	0,137
11	-0,010	0,060	0,256	49	-0,016	-0,561	-0,115
12	0,332	-0,537	0,258	50	0,192	0,065	0,183
13	0,280	0,224	0,086	51	0,283	0,013	-0,018
14	0,033	-0,081	0,127	52	0,557	0,092	0,121
15	0,286	0,206	-0,032	53	-0,096	-0,160	-0,014
16	0,065	-0,118	0,020	54	-0,160	-0,781	-0,150
17	-0,177	-0,243	0,150	55	0,507	-0,275	-0,019
18	-0,265	-0,004	0,052	56	0,517	0,160	-0,544
19	0,686	0,179	-0,009	57	-0,214	0,298	0,755
20	0,599	0,194	0,164	58	0,124	-0,054	0,986
21	0,725	0,396	0,089	59	0,065	0,855	0,120
22	0,572	-0,273	-0,242	60	-0,080	0,097	0,639
23	0,511	0,104	-0,094	61	0,010	0,593	0,482
24	0,960	0,156	-0,232	62	-0,098	-0,194	0,379
25	0,325	0,367	0,359	63	-0,276	-0,045	0,242
26	-0,155	-0,291	-0,458	64	0,049	0,895	-0,140
27	0,489	0,064	0,201	65	-0,050	0,919	0,152
28	0,395	0,465	0,199	66	0,656	0,355	-0,180
29	0,447	0,194	-0,131	67	-0,005	0,100	-0,106
30	0,777	0,174	0,097	68	-0,250	0,009	0,065
31	0,295	-0,130	-0,052	69	-0,236	-0,464	0,264
32	0,630	0,250	0,041	70	0,096	0,460	0,391
33	0,761	-0,166	0,085	71	0,084	-0,191	0,635
34	0,351	0,003	0,086	72	-0,525	-0,322	-0,097
35	0,492	-0,114	0,166	73	0,065	0,995	-0,055
36	0,275	0,297	-0,136	74	0,012	0,093	0,948

Item	Dimensões			Item	Dimensões		
	1	2	3		1	2	3
37	0,575	-0,347	-0,052	75	-0,126	-0,083	0,494
38	0,698	0,276	-0,052				

Verifica-se que a maioria dos itens apresenta cargas superiores a 0,300, em alguma dimensão, que é considerado por alguns autores (JOHNSON, WICHERN 2007, HAIR et al., 2009) um valor mínimo para que se possa considerar o item na interpretação da dimensão. Itens com carga fatorial inferior a 0,300 em todas as dimensões foram assumidos como itens pouco informativos para o construto, sendo assim, foram eliminados. Esta ação resultou na retirada de 18 itens. A nova análise do conjunto, após esta primeira retirada, mostrou mais dois itens com carga fatorial, em módulo, inferior a 0,300 em todas as dimensões, foram os itens 01 e 42, os quais também foram eliminados. Foram realizadas cinco novas análises até extrair todos os itens remanescentes com carga fatorial inferior a 0,300, desta forma, foram eliminados ao todo 31 itens. A descrição destes itens eliminados pode ser visualizada no Quadro 4.

Quadro 4- Descrição dos 31 itens eliminados.

Item eliminado	Descrição
01	Ao abrir a homepage há abertura de janelas pop-up?
02	A homepage apresenta as principais áreas de conteúdo do site (navegação)?
04	A homepage apresenta um resumo das promoções mais importantes?
07	Os menus estão organizados em ordem alfabética?
09	Para navegação, existe uma trilha à esquerda ou links na parte superior?
11	O título da janela (browser) lista o nome do site em primeiro lugar?
13	As páginas apresentam consistência visual, ou seja, mantém sempre a mesma aparência visual?
14	O site possui a opção de acesso em outras línguas?
15	As palavras clicáveis (coloridas ou sublinhadas) apresentam uma forma diferenciada quando são selecionadas?
16	Os rótulos de campos começam com uma letra maiúscula, e as letras restantes são minúsculas?

<b>Item eliminado</b>	<b>Descrição</b>
17	Os títulos estão alinhados a esquerda?
18	Os parágrafos de texto são separados?
20	Os títulos de telas, janelas e caixas de diálogo estão no alto, centrados ou justificados à esquerda?
26	Há cores diferentes para links já visitados?
31	Listas longas apresentam indicadores de continuação, de quantidade de itens e de páginas?
34	É possível ampliar as fotos dos produtos para visualizar detalhes?
36	Os grupos de botões de comando estão dispostos em coluna e à direita, ou em linha e abaixo dos objetos aos quais estão associados?
39	Existe imagens de fundo nos textos?
41	Existe a opção de compartilhar a página nas redes sociais?
42	Existe a opção de ajuda on-line?
44	A empresa oferece algum serviço grátis, como por exemplo frete grátis?
49	Os campos destinados a formulários de preenchimento pelo usuário estão agrupados linearmente, evitando espaços desnecessários?
50	No preenchimento de um formulário, as opções que não são válidas ou não estão disponíveis estão visivelmente desativadas, prevendo assim erros?
51	No preenchimento de um formulário, é informado a forma de preenchimento?
53	É possível efetivar uma compra sem realizar um cadastro (que inclua login e senha)?
54	O sistema emite sinais sonoros quando ocorrem problemas na entrada de dados?
62	A apresentação visual do texto e imagens de texto tem uma relação de contraste de, no mínimo, 7:1?
63	Toda a funcionalidade do conteúdo é operável através de uma interface de teclado sem a necessidade de qualquer espaço de tempo entre cada digitação individual, exceto quando a função subjacente requer entrada de dados que dependa da cadeia de movimento do usuário e não apenas dos pontos finais?
67	As páginas Web têm títulos que descrevem o tópico ou a finalidade?
68	A finalidade de cada link pode ser determinada a partir apenas



<b>Item eliminado</b>	<b>Descrição</b>
	do texto do link ou a partir do texto do link juntamente com o respectivo contexto do link determinado de forma programática, exceto quando a finalidade do link for ambígua para os usuários em geral?
72	Os componentes que têm a mesma funcionalidade num conjunto de páginas Web são identificados de forma consistente?

Do ponto de vista estatístico, estes itens não estão correlacionados com os demais itens, implicando que, se o objetivo é medir qualidade de websites, estes itens teoricamente não estão associados a este objetivo. Do ponto de vista prático pode ser verificado que, no caso do item 01, conforme argumentado por Tezza (2009), no qual este mesmo item foi avaliado e descartado em um construto unidimensional, o mesmo pode não caracterizar um item cumulativo. Ou seja, a possibilidade de abertura de pop-up durante a interação com o website é vista na literatura como ruim para a qualidade e desnorteador do usuário (STOREY et al., 2002; PETRE et al. 2006; NIELSEN 2006). No entanto, esta característica é complexa porque envolve questões comerciais e de maturidade tecnológica. O fato de exigir certo grau de maturidade tecnológica pode indicar que esta característica avaliada puramente como existe ou não existe a abertura de janelas pop-up, pode não avaliar de forma linear ou cumulativa com a qualidade de um websites. Por exemplo, websites que não possuem maturidade tecnológica possuem uma maior probabilidade de não apresentar pop-up, entretanto nas demais características de qualidade estes websites possuem maior probabilidade de não apresentarem, o que pode caracterizar teoricamente uma inversão prática. Esta mesma lógica justifica o item 04 relacionado com a apresentação de um resumo das promoções mais importantes.

O mesmo ocorre com o item 14 o qual é descrito como opção de acesso em outras línguas. Aparentemente esta característica parece estar relacionada com acessibilidade e websites de conteúdo geral. Entretanto, em websites comerciais, possui uma ligação direta com os objetivos estratégicos e comerciais da empresa, bem como a característica do produto comercializado.

Os itens 02, 09, 20, 36 e 39 são itens que teoricamente possuem uma ligação direta com qualidade de websites comerciais. Entretanto, uma verificação das respostas a estes, com base nos 441 webistes analisados, mostra que se trata de características uniformes, ou seja, a quase totalidade dos websites possui esta característica. Este comportamento dificulta o

tratamento estatístico, uma vez que este é baseado em padrões de respostas e suas relações. Esta constatação mostra que características como apresentar as principais áreas de conteúdo do site, configurar trilhas de navegação a esquerda ou na parte superior, e ainda agrupar os botões de comando a direita ou na linha de baixo do objeto associado, são características mais do que básicas em uma interface, fazendo assim parte dos modelos mentais dos usuários e dos desenvolvedores de websites comerciais. No caso do item 39, descrito como “não existe imagens de fundo nos textos”, segundo Nielsen e Loranger (2006), imagens ou cores de fundo nos textos reduzem a legibilidade e consequentemente interferem na qualidade da apresentação da informação. A análise mostrou que raríssimos websites comerciais arriscam inserir figuras de fundo, possivelmente se estivesse tratando de websites não comerciais esta característica sofresse alguma alteração, pois nestes o objetivo estético tende a se sobressair ao técnico.

Situação inversa aos itens de baixíssima dificuldade observa-se nos itens 26 e 53, descritos como “há cores diferentes para links visitados e não visitados” e “é possível efetivar uma compra sem realizar um cadastro (que inclua login e senha)” respectivamente. Nielsen e Lorange (2006), comparando seus estudos em Nielsen (2000) constataam que apesar de nortear a navegação dos usuários, a mudança de cor dos links já visitados, é ignorada pela grande maioria dos websites comerciais, o que pode ser constatado no presente estudo. Websites de busca como o Google.com e a maioria dos noticiários adotam esta prática, entretanto, a amostra dos 441 websites analisados não possuem esta característica e quando a possuem, isso não reflete necessariamente em um aumento conjunto das demais características de qualidade da interface. O mesmo ocorre com o item 53, com a peculiaridade de que neste caso, exigir um cadastro pode não significar uma redução de qualidade da interface, mas sim uma intenção de personalizar a interação e facilitar as interações futuras.

A característica “existe a opção de ajuda on-line” descrita no item 42, é defendida no âmbito geral como útil principalmente para usuários novatos, representando assim um acréscimo positivo na experiência deste (HHS, 2006). Cybis (2007) argumenta que toda interface deve apresentar algum tipo de ajuda contextual e específica. Entretanto, esta característica parece não estar relacionada diretamente com aumento da qualidade de um website comercial. Tezza (2009), considerando esta característica associada a um construto unidimensional para avaliar usabilidade chegou à conclusão estatística semelhante. Isto pode ser justificado pelo fato de que a opção “ajuda” não tem função direta na tarefa de compra em um site comercial, ela é utilizada principalmente quando alguma das características básicas de usabilidade falha.

O item 41 (existe a opção de compartilhar a página nas redes sociais?) representa uma alternativa comercial e interativa relativamente nova, e praticamente sem custo ao website, no entanto, os resultados da pesquisa mostraram que esta característica necessariamente não aumenta a qualidade de um website comercial. O mesmo ocorre com o item 44 (a empresa oferece algum serviço grátis, como por exemplo frete grátis?) o qual avalia se o website oferece ou não algum benefício financeiro para o usuário, no entanto, neste caso há uma contrapartida financeira por parte do website. Este item é encontrado com frequência na literatura (YANG; JUN, 2002; LONG; McMELLON, 2004; COLLIER; BIENSTOCK, 2006) associado a percepção do usuário, o que em termos de análise do sistema não foi verificado consistência com a qualidade.

Os demais itens eliminados, teoricamente, possuem algum tipo de relação com a qualidade de um website comercial, entretanto, a análise dos 441 websites não demonstrou uma relação estatística consistente, isso pode ter sido gerado pela formulação do item ou mesmo pelo tipo diferenciado de análise que ao invés de avaliar a percepção do usuário, avalia as condições do sistema.

Após a eliminação destes itens com baixa carga fatorial deu-se sequência à análise dos itens restantes que constituirão as modelagens subsequentes.

#### 4.3 VALIDAÇÃO DAS DIMENSÕES

O procedimento utilizado para fazer a validação das dimensões do construto foi o empírico, ou seja, a definição das dimensões baseou-se na carga fatorial do item e a caracterização dos conceitos centrais de cada item.

Antes de iniciar a análise das dimensões, realizou-se uma nova análise da dimensionalidade, nesta etapa apenas com os 44 itens restantes.

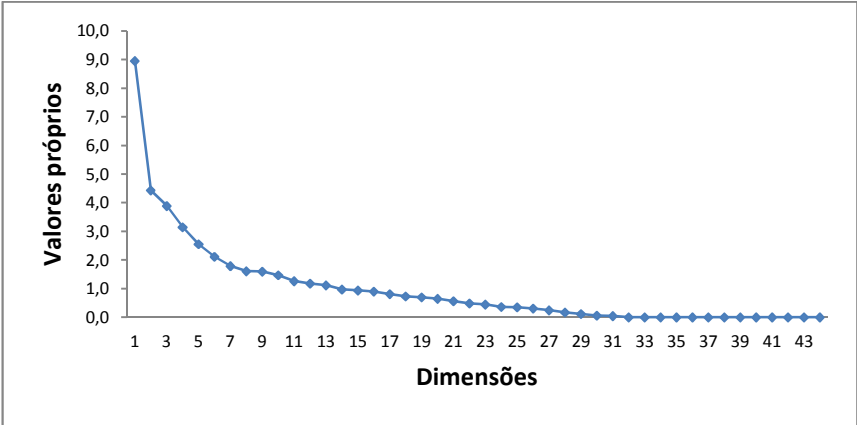
A análise de informação restrita, realizada por meio da matriz de correlação tetracórica sugere a existência de no mínimo cinco dimensões, assumindo o critério de pelo menos 50% da variância comum para o conjunto de fatores, conforme descrito na seção 4.1. A Tabela 6 mostra os primeiros 10 autovalores considerando os 44 itens e a Figura 12 mostra o gráfico considerando os 44 autovalores.

Tabela 6 - Valores próprios da matriz de correlação tetracórica dos primeiros 10 autovalores do conjunto de 44 itens.

Dimensão	1	2	3	4	5	6	7	8	9	10
Autovalor	8,95	4,44	3,89	3,15	2,55	2,12	1,79	1,61	1,60	1,47
Proporção acumulada da variação explicada	20,34	30,42	39,26	46,41	52,22	57,03	61,09	64,74	68,38	71,72

Verifica-se na Tabela 6 que, com 5 dimensões, a proporção acumulada da variação explica ou variância comum é de 52,22, sugerindo desta forma cinco dimensões para este construto.

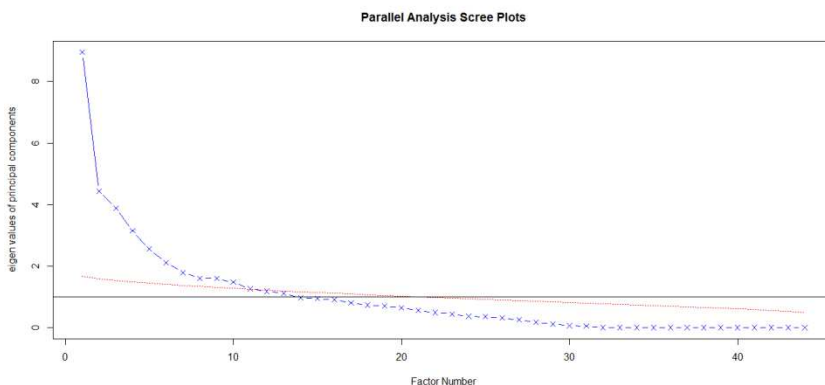
Figura 12: Valores próprios da matriz de correlação tetracórica considerando os 44 itens.



Na Figura 12, é possível visualizar o scree-plot referente ao conjunto de autovalores considerando os 44 itens. A interpretação deste scree-plot pode ser feita com base no ponto de curvatura da curva, ou seja, espera-se que o número de dimensões do conjunto de dados esteja próximo deste ponto de curvatura, que neste caso aparentemente está posicionado próximo de 5 dimensões.

A análise paralela sugere a existência de 14 dimensões, como pode ser visualizado no gráfico da Figura 13.

Figura 13 - Análise paralela dos 44 itens restantes.



Já a análise feita pelo método de informação plena mostrou que a primeira dimensão explica apenas 10,41% da variação total do conjunto de 44 itens, o que caracteriza, assim como a análise de informação restrita e plena dos 75 itens iniciais, um construto com mais de uma dimensão. A comparação dos modelos de uma, duas, três e quatro dimensões por meio dos critérios de informação BIC e AIC sugerida por Bartolucci et al (2012) mostrou evidências da existência de três dimensões, assumindo como na seção 4.1 o critério AIC, como pode ser visualizado na tabela 7.

Tabela 7 - Seleção do melhor modelo com base nos critérios de informação AIC e BIC.

Modelo	-2 Log da máxima verossimilhança	AIC	BIC
Mod1	-6849,03	14068,05	14824,53
Mod2	-6617,70	13691,87	14623,69
Mod3	-6531,43	<b>13602,87</b>	14706,91
Mod4	-6493,57	13609,14	14880,84

Desta forma, testaram-se empiricamente as estruturas com três, quatro e cinco dimensões, com a finalidade de verificar qual destas possui um melhor ajuste para o construto em questão.

Na estrutura contendo três dimensões, as correlações entre estas dimensões foram relativamente baixas, como mostra a Tabela 8, o que permite a utilização de rotação ortogonal.

Tabela 8 - Correlação linear dos fatores assumindo três dimensões dos 44 itens.

	F_1	F_2	F_3
F_1	1,000	0,121	-0,046
F_2	0,121	1,000	0,208
F_3	-0,046	0,208	1,000

A Tabela 9 traz a estrutura tridimensional assumindo a rotação oblíqua promax e a rotação ortogonal varimax.

Tabela 9 - Análise fatorial utilizando o método da informação plena, assumindo rotação oblíqua promax e ortogonal varimax aos 44 itens em três dimensões.

Item	Rotação promax			Rotação varimax			Comunalidade
	F_1	F_2	F_3	F_1	F_2	F_3	
03	<b>0,459</b>	0,380	0,156	<b>0,493</b>	0,408	0,173	0,440
05	-0,434	<b>0,561</b>	0,020	-0,376	<b>0,552</b>	0,061	0,450
06	0,377	<b>0,573</b>	0,218	0,429	<b>0,608</b>	0,247	0,614
08	<b>-0,572</b>	0,331	0,219	<b>-0,545</b>	0,351	0,245	0,480
10	<b>0,432</b>	0,057	-0,278	<b>0,448</b>	0,023	-0,276	0,278
12	0,094	<b>0,557</b>	-0,213	0,160	<b>0,523</b>	-0,176	0,330
19	<b>0,697</b>	-0,076	0,014	<b>0,689</b>	-0,061	0,000	0,478
21	<b>0,765</b>	-0,289	-0,104	<b>0,739</b>	-0,289	-0,131	0,646
22	<b>0,475</b>	0,132	0,285	<b>0,479</b>	0,182	0,284	0,343
23	<b>0,484</b>	-0,007	0,171	<b>0,477</b>	0,027	0,163	0,255
24	<b>0,985</b>	-0,319	0,157	<b>0,946</b>	-0,276	0,122	0,986
25	0,087	-0,056	<b>-0,483</b>	0,098	-0,126	<b>-0,482</b>	0,258
27	<b>0,500</b>	0,117	-0,168	<b>0,518</b>	0,100	-0,165	0,306
28	<b>0,470</b>	-0,408	-0,260	<b>0,437</b>	-0,436	-0,289	0,464
29	<b>0,578</b>	-0,158	0,165	<b>0,556</b>	-0,123	0,146	0,345
30	<b>0,774</b>	-0,031	-0,101	<b>0,774</b>	-0,032	-0,112	0,613
32	<b>0,658</b>	-0,101	0,042	<b>0,646</b>	-0,082	0,027	0,424
33	<b>0,642</b>	0,317	-0,118	<b>0,679</b>	0,308	-0,105	0,567
35	<b>0,379</b>	0,119	-0,076	<b>0,394</b>	0,114	-0,072	0,173
37	<b>0,406</b>	0,400	0,101	<b>0,444</b>	0,419	0,120	0,387
38	<b>0,752</b>	-0,232	0,056	<b>0,726</b>	-0,209	0,030	0,571
40	<b>0,430</b>	0,306	-0,083	<b>0,465</b>	0,299	-0,067	0,310

Item	Rotação promax			Rotação varimax			Comunalidade
	F_1	F_2	F_3	F_1	F_2	F_3	
43	0,137	<b>0,400</b>	-0,154	0,185	<b>0,376</b>	-0,128	0,192
45	<b>0,572</b>	0,410	-0,058	<b>0,617</b>	0,408	-0,038	0,549
46	<b>0,626</b>	-0,322	-0,244	<b>0,601</b>	-0,346	-0,270	0,554
47	0,353	<b>0,468</b>	0,167	0,396	<b>0,495</b>	0,191	0,439
48	<b>0,709</b>	0,238	-0,103	<b>0,738</b>	0,233	-0,096	0,608
52	<b>0,526</b>	-0,094	-0,089	<b>0,519</b>	-0,097	-0,100	0,289
55	0,349	<b>0,401</b>	0,070	0,388	<b>0,415</b>	0,091	0,331
56	<b>0,674</b>	0,199	0,240	<b>0,686</b>	0,245	0,241	0,589
57	-0,209	0,083	<b>-0,804</b>	-0,171	-0,041	<b>-0,787</b>	0,650
58	-0,054	0,231	<b>-1,019</b>	0,007	0,076	<b>-0,992</b>	0,989
59	0,217	<b>-0,756</b>	-0,187	0,145	<b>-0,774</b>	-0,236	0,676
60	-0,021	-0,116	<b>-0,751</b>	-0,006	-0,227	<b>-0,749</b>	0,613
61	0,166	-0,387	<b>-0,492</b>	0,144	-0,455	<b>-0,513</b>	0,491
64	0,225	<b>-0,855</b>	0,218	0,128	<b>-0,812</b>	0,157	0,700
65	0,298	<b>-0,955</b>	-0,106	0,203	<b>-0,959</b>	-0,170	0,989
66	<b>0,728</b>	-0,382	0,273	<b>0,679</b>	-0,326	0,236	0,622
69	-0,375	<b>0,585</b>	-0,256	-0,305	<b>0,536</b>	-0,211	0,425
70	0,166	-0,268	<b>-0,451</b>	0,154	-0,330	<b>-0,465</b>	0,349
71	0,049	0,455	<b>-0,524</b>	0,115	0,374	<b>-0,489</b>	0,392
73	0,092	<b>-1,017</b>	0,352	-0,027	<b>-0,956</b>	0,281	0,992
74	-0,069	0,221	<b>-0,992</b>	-0,010	0,070	<b>-0,966</b>	0,937
75	-0,180	0,334	<b>-0,529</b>	-0,126	0,249	<b>-0,499</b>	0,327

Observa-se na Tabela 9 que a estrutura de carregamento dos fatores é idêntica tanto para a rotação oblíqua quanto para a rotação ortogonal, em destaque, a dimensão em que o item possui maior carga fatorial. Nesta estrutura, verifica-se também que na primeira dimensão há uma predominância de itens com carregamento positivo, apenas com exceção do item 08. O mesmo ocorre na dimensão 2. No entanto, na dimensão 2, quatro dos onze itens possuem carga com sinal contrário, o que estatisticamente significa uma relação inversa nesta dimensão. A análise com base no conteúdo de cada item nas referidas dimensões não mostrou clareza na

primeira e na segunda dimensão, pois ambas misturam conceitos de navegação, apresentação da informação e interação do usuário. No caso da primeira dimensão, nota-se a existência de itens relacionados ao conceito de navegação (por exemplo, itens 21, 24, 46, 52), apresentação da informação (por exemplo, itens 03, 27, 33, 35, 37, 40, 45, 48, 56) e interatividade (por exemplo, itens 28, 29, 38). Comportamento semelhante pode ser visualizado na segunda dimensão, por exemplo, os itens 12, 43, 47, 55 e 59 estão associados ao conceito de apresentação da informação, apesar do item 59 apresentar carga negativa, o que não é esperado uma vez que este item está relacionado com a utilização de cores para apresentação de algum tipo de informação. Os itens 64, 65 e 73, que também apresentaram carga negativa nesta dimensão, estão associados com o controle do usuário e o item 69 descrito como “os cabeçalhos e as etiquetas descrevem o tópico ou a finalidade” possui características tanto de informação quanto de acessibilidade do sistema. Esta última característica, associada à acessibilidade do sistema é predominante na terceira dimensão, a qual conta com os itens 57, 58, 60, 61, 70, 71, 74 e 75 com exceção do item 25 “O site permite navegação em suas páginas em apenas uma janela, ou seja, não há abertura de novas janelas em meio a navegação?” o qual avalia a qualidade da navegação. Sendo assim, esta estrutura tridimensional não consegue diferenciar claramente os conceitos de qualidade em cada uma das três dimensões, independente do tipo de rotação.

Na sequência, a mesma análise foi feita considerando 4 dimensões. A Tabela 10 mostra a análise de correlação das dimensões.

Tabela 10: Correlação dos fatores assumindo quatro dimensões dos 44 itens.

	F_1	F_2	F_3	F_4
F_1	1,000	0,176	-0,168	-0,387
F_2	0,176	1,000	-0,165	-0,161
F_3	-0,168	-0,165	1,000	-0,148
F_4	-0,387	-0,161	-0,148	1,000

Nota-se que assim como a estrutura tridimensional (Tabela 8), as dimensões apresentam baixa correlação. Desta forma, optou-se por avaliar apenas a estrutura rotacionada ortogonalmente.

A Tabela 11 mostra a distribuição das cargas fatoriais de cada item nas quatro dimensões, utilizando a rotação ortogonal varimax.



Tabela 11 - Análise fatorial utilizando o método da informação plena, assumindo rotação ortogonal varimax aos 44 itens em quatro dimensões.

Item	F_1	F_2	F_3	F_4	Comunalidade
03	0,130	<b>0,660</b>	0,030	-0,020	0,454
05	-0,160	0,140	-0,200	<b>-0,440</b>	0,279
06	0,390	<b>0,680</b>	0,110	-0,090	0,635
08	-0,060	0,070	-0,190	<b>-0,570</b>	0,370
10	-0,050	0,300	0,030	<b>0,530</b>	0,374
12	-0,060	<b>0,450</b>	0,380	-0,300	0,441
19	0,290	0,390	0,220	<b>0,400</b>	0,445
21	<b>0,450</b>	0,280	<b>0,590</b>	0,300	0,719
22	0,260	<b>0,380</b>	0,270	-0,280	0,363
23	0,240	<b>0,480</b>	0,260	0,270	0,429
24	<b>0,520</b>	0,250	<b>0,750</b>	0,320	0,998
25	0,110	-0,030	<b>0,820</b>	-0,280	0,764
27	0,030	0,300	-0,080	<b>0,480</b>	0,328
28	0,060	0,020	0,180	<b>0,660</b>	0,472
29	0,220	0,170	-0,170	<b>0,530</b>	0,387
30	0,290	0,390	0,300	<b>0,500</b>	0,576
32	0,290	<b>0,310</b>	<b>0,310</b>	0,290	0,360
33	0,280	<b>0,710</b>	0,380	0,000	0,727
35	0,050	<b>0,330</b>	0,050	<b>0,330</b>	0,223
37	0,270	<b>0,620</b>	-0,040	0,110	0,471
38	<b>0,400</b>	0,200	0,290	<b>0,460</b>	0,496
40	0,070	<b>0,450</b>	0,030	0,240	0,266
43	0,000	<b>0,520</b>	-0,090	0,080	0,285
45	0,130	<b>0,520</b>	0,020	0,330	0,397
46	0,210	0,200	<b>0,530</b>	0,410	0,533
47	0,240	0,430	<b>0,470</b>	-0,240	0,521
48	0,330	<b>0,600</b>	0,290	0,250	0,616
52	0,280	0,220	<b>0,440</b>	0,220	0,369
55	0,210	0,100	<b>0,540</b>	-0,200	0,386
56	<b>0,580</b>	<b>0,550</b>	0,280	0,100	0,727

Item	F_1	F_2	F_3	F_4	Comunalidade
57	<b>-0,710</b>	-0,050	0,400	-0,010	0,667
58	<b>-0,960</b>	0,160	0,130	0,190	1,000
59	0,150	<b>-0,460</b>	<b>0,520</b>	0,270	0,577
60	<b>-0,720</b>	-0,120	0,010	0,510	0,793
61	-0,340	-0,180	<b>0,430</b>	0,280	0,411
64	0,220	<b>-0,600</b>	<b>0,550</b>	0,100	0,721
65	0,060	<b>-0,590</b>	<b>0,670</b>	0,230	0,854
66	<b>0,440</b>	-0,060	0,320	<b>0,400</b>	0,460
69	<b>-0,430</b>	0,160	-0,230	-0,200	0,303
70	-0,180	-0,050	<b>0,420</b>	0,220	0,260
71	<b>-0,490</b>	0,220	0,220	0,120	0,351
73	0,180	<b>-0,740</b>	<b>0,520</b>	-0,050	0,853
74	<b>-0,880</b>	0,120	0,260	0,130	0,873
75	<b>-0,410</b>	0,150	0,170	-0,120	0,234

A segunda dimensão traz itens relacionados diretamente com apresentação da informação, praticamente todos os itens com maior carga fatorial nesta dimensão remetem a algum conceito de organização ou apresentação da informação, por exemplo, o item 03 trata da apresentação da informação na homepage, que é porta de entrada de um website. O item 06, assim como os itens 12 e 45, trata de confiabilidade na apresentação da informação. Os itens 22, 23 e 32 referem-se à apresentação da informação durante a navegação, por exemplo, o item 22 descrito como “quando há rolagem, não existem elementos de design (na tela inicial) que pareçam como marcadores de final de página” remete a maneira como a informação ou o conteúdo do websites esta organizado, inserir elementos de design que pareçam como marcadores de final de página pode confundir o usuário durante a navegação sugerindo a ele que não há mais informação na rolagem da página. Segundo Nielsen e Loranger (2006), geralmente o usuário não toma atenção à barra de rolagem que fica no canto esquerdo da tela, podendo então não encontrar informações relevantes para seu objetivo de compra. Já o item 23 descrito como “o logotipo da empresa está no canto superior esquerdo em todas as páginas do site” está associado com a localização do usuário baseado na apresentação da informação do logotipo da empresa. Uma das mais fundamentais características de qualidade de

websites comerciais é a apresentação de informações referentes ao produto (GOENARIS; DIMITRIADIS, 2003; LIU et al., 2009). Esta característica é abordada nos itens 33, 35, 37, 43 e 48, os quais pertencem à segunda dimensão, segundo esta modelagem. Os itens 64 e 73 também apresentaram maior carga no segundo fator, entretanto, apresentaram carga com sinal contrário ao da maioria dos itens. Isto significa, segundo Hair et al., (2009), que as características dos itens variam juntas mas, se movem em sentidos opostos um em relação ao outro. Ou seja, o item 64, o qual trata de controle de tempo por parte do usuário pode ser justificado como pertencente a dimensão de apresentação da informação porque geralmente estas situações práticas envolvem apresentação da informação. No entanto, a possibilidade de interação do usuário com esta forma de apresentação e/ou interrupção do mesmo enquanto acessa a informação pode caracterizar um sentido oposto. Ainda na análise transversal do item 64, nota-se que tal item apresenta uma característica fortemente cruzada, termo utilizado por Hair et al. (2009) para itens que possuem cargas elevadas sobre dois ou mais fatores. Neste caso, como veremos adiante o fator 3, no qual o item possui carga elevada também, relaciona-se com interatividade. O item 73 tem um comportamento semelhante, este item trata da descrição automática o usuário de possíveis erros. O sentido negativo do item pode estar associado com o fato de que o texto apresentado ao usuário de forma automática pode não ser informativo ou explicativo o que decorre em possível desinformação.

A terceira dimensão, assim como a segunda apresentou todos os itens com maior carga nesta dimensão no mesmo sentido, ou seja, todos positivos, com exceção do item 73. Analisando-se os conceitos envolvidos em cada item característico desta dimensão, verifica-se que a maioria possui conceito adjacente ao controle por parte do usuário ou interação deste com o sistema. O item 21 traz uma das principais alternativas de interação do usuário com o website que é a apresentação da ferramenta de busca, permitindo assim que o usuário procure em qualquer subpágina do website o produto de interesse. O item 24 permite que o usuário possa voltar diretamente para a página inicial clicando em um link direto, dando-lhe a sensação de controle defendida por Cybis et al. (2007), Bastien e Scapin (1993) e Padovani e Moura (2008). O item 25, por sua vez trata da possibilidade do usuário inserir *feedback* sobre um produto o que reflete uma alternativa de interação e confiabilidade. O item 47 discute alternativa de pagamento, além do cartão de crédito o que pode significar uma flexibilidade entre usuário e vendedor, como é argumentado por Yoo e Donthu (2001) e Hung e McQueen (2004). O item 55 trata da interação/resposta do sistema no preenchimento de formulários. O item 61

trata da possibilidade de ampliar o texto em até 200% o que pode ser útil para usuários com algum tipo de deficiência visual. Os itens 64 e 65 incluem a característica de controle da interação por parte do usuário, que é abordada nestes itens como a possibilidade do usuário pausar uma alternativa multimídia ou desligar o limite de tempo pré definido pelo sistema ou até mesmo ajustá-lo, estes itens são requisitos abordados nas recomendação da WCAG 2.0 (W3C, 2008). Ding et al., (2011) também particulariza uma dimensão associada à percepção de controle por parte do usuário.

A primeira dimensão assemelha-se à terceira dimensão visualizada no modelo tridimensional, com a exceção dos itens 25, 61, 69 e 70. Desta forma, esta dimensão agrupa itens com conteúdo voltado para acessibilidade e utilização do sistema, ou seja, verificação de possíveis entraves no sistema que possam impedir, confundir ou dificultar o acesso dos usuários, particularmente usuários que possuam alguma limitação física ou cognitiva. Nota-se que a maioria dos itens desta dimensão pertencem ao conjunto de recomendações de acessibilidade WCAG 2.0 (W3C, 2008). Os itens 57, 58 e 75 possuem como conceito central a possibilidade de transformar qualquer conteúdo do website em conteúdo textual e consequentemente possível de ser lido por programas próprios para usuários com deficiência visual, as chamadas tecnologias assistidas. Os itens 69, 71 e 74 abordam características do sistema que facilitam a navegação e a procura de informação. Nota-se particularmente que esta dimensão possui, diferentemente das demais, um carregamento no sentido negativo. Segundo Hair et al. (2009), pelo fato de se estar considerando rotação ortogonal, o sentido das cargas só deve ser analisado dentro de cada fator uma vez que a rotação ortogonal assume que não há correlação entre os fatores. Desta forma, os itens 21, 24, 38, 56 e 66, os quais possuem carga positiva no fator 1 podem ser interpretadas como correlacionadas a acessibilidade da forma como caracteriza os demais itens com carga negativa neste fator no entanto, assumindo sentido oposto. O mesmo acontece com os itens 5 e 8 na dimensão 4 e com o item 73 na dimensão 2. Uma carga negativa indicando que o fator apresenta uma direção distinta dos demais, isto pode-se justificar pelo fato de a característica de acessibilidade ser muito mais ligada ao sistema computacional que permite ou não o acesso de usuários com algum tipo de dificuldade, enquanto que os demais fatores relacionam-se com características ligadas a apresentação gráfica e navegação.

A quarta dimensão, ao contrário das demais, apresenta dois itens com carga em sentido contrário aos da maioria. Esta peculiaridade pode ser justificada pela própria característica dos itens neste contexto. Ou seja, considerando que todos os itens desta dimensão estão relacionados

diretamente com o conceito de orientação do usuário durante a navegação, o item 05 “o website não possui imagens em movimento que podem distrair o usuário” pode representar uma possibilidade de orientação invertida, considerando que esta característica, diferentemente da abordada pelo item 01 (abertura de janelas pop up) não representa a abertura de novas janelas o que de fato pode prejudicar a navegação do usuário. Imagens em movimento dependendo dos objetivos podem ser classificadas como orientação, dimensão evidenciada por autores como Yang et al., (2005), Chang e Chen (2009) ou divertimento, dimensão evidenciada por autores como Bauer et al., (2006) e Nusair e Kandampully (2008) além disso, autores como Fassnacht e Koese (2006) incluem em suas análises o fator benefício emocional, que está associado à apresentação de objetos divertidos e que possam interagir emocionalmente com o usuário. Evidentemente que a análise deste item é mais complexa do que parece, pois aparentemente trata do envolvimento do usuário durante a navegação, o que não foi objeto de análise direta neste trabalho. Entretanto, o comportamento estatístico deste item indicou uma tendência no sentido de não ser necessariamente uma característica que reduza a qualidade de um website, pelo menos não da maneira em que foi avaliada neste trabalho.

O item 08 trata de uma característica não muito clara na literatura quanto a seu impacto na qualidade de um website “o website não possui menu em cascata”. Interfaces envolvendo menus em cascata segundo alguns autores deve ser evitada, pois esta função tende a dificultar a navegação de usuários que tenham algum tipo de limitação motora (NIELSEN ; TAHIR, 2002; NIELSEN; LORANGER, 2006). No entanto, outros autores vêem nesta característica a possibilidade de agrupamento de informação sem reduzir a quantidade de informação além de fornecer mais opções de navegação orientada (ROSENFELD; MORVILLE, 1998). De qualquer forma, esta característica está associada à navegação e orientação do usuário na interface, tendo sua navegação prejudicada ou ampliada dependendo o ponto de vista. Sendo assim, tal característica possui o mesmo comportamento da característica abordada no item 05, verificada aqui, na análise de sistemas como sendo uma característica boa ter menu em cascata e isto está associado à orientação/navegação.

Os demais itens característicos desta dimensão estão ligados diretamente com a navegação e orientação do usuário, como por exemplo, os itens 19 e 38; os quais referem-se a orientação do usuário em meio à navegação no website. O item 19 “as palavras aparentemente clicáveis são de fato clicáveis” é um item claramente ligado com a navegabilidade de um website. Segundo Nielsen e Loranger (2006), uma boa navegação é previsível e faz com que as pessoas se sintam confortáveis em explorar o

website. Padovani e Moura (2008) chamam esta característica de navegação visualmente clara e defendem que o uso de mecanismos de diferenciação e codificação visual pode tornar o sistema mais claro. Uma característica que pode desorientar o usuário é a abertura repentina de novas páginas durante a navegação que é verificada no item 38. O item 27 “existe uma lista de perguntas frequentes” pode ser, segundo Hernández et al, (2007), uma possibilidade de auxílio do sistema para o usuário no cumprimento de seus objetivos ou simplesmente um alternativa de informação. Os itens 28, 29 e 30 que estão associados diretamente com a ferramenta de busca as quais possuem como objetivo conduzir e interagir com usuário na interface para cumprir sua tarefa de compra ou pesquisa. O item 10 “existe agrupamento de subcategorias” pode ser tomado também com uma forma de apresentação de informação.

O Quadro 5 mostra um resumo das dimensões e os respectivos itens de maior carregamento nestas segundo a análise fatorial.

Quadro 5 - Classificação dos itens conforme a dimensão, segundo a análise fatorial.

<b>Dimensão</b>	<b>Itens</b>
Acessibilidade/utilização do sistema	56 <sup>*</sup> , 57, 58, 60, 66 <sup>*</sup> , 69, 71, 74, 75
Apresentação da informação	03, 06, 12, 22, 23, 32, 33, 35, 37, 40, 43, 45, 48, 56, 64 <sup>*</sup> , 73 <sup>*</sup>
Controle por parte do usuário/interatividade	21, 24, 25, 32, 46, 47, 52, 55, 59, 61, 64, 65, 70
Orientação do usuário/navegabilidade	05 <sup>*</sup> , 08 <sup>*</sup> , 10, 19, 27, 28, 29, 30, 35, 38, 66

<sup>\*</sup>Itens com carregamento contrário a maioria dos itens da dimensão

Realizando-se a mesma análise, agora considerando cinco dimensões, verifica-se que a correlação das dimensões sofreram algumas modificações com a inserção de uma nova dimensão, aumentando algumas relações e diminuindo outras. A Tabela 12 mostra a análise de correlação das dimensões.

Tabela 12 - Correlação dos fatores assumindo 5 dimensões dos 44 itens.

	F_1	F_2	F_3	F_4	F_5
F_1	1,000	-0,060	0,000	0,004	-0,245
F_2	-0,060	1,000	-0,048	0,509	-0,447
F_3	0,000	-0,048	1,000	0,078	0,157
F_4	0,004	0,509	0,078	1,000	-0,284
F_5	-0,245	-0,447	0,157	-0,284	1,000

Considerando a existência de uma correlação de 0,509 entre as dimensões 4 e 2, verificou-se as estruturas fatoriais tanto na rotação ortogonal quanto na oblíqua. A Tabela 13 mostra a distribuição das cargas fatoriais de cada item nas cinco dimensões, utilizando a rotação oblíqua promax e a rotação ortogonal varimax.

Tabela 13 - Análise fatorial utilizando o método da informação plena, assumindo rotação ortogonal varimax e oblíqua promax aos 44 itens em cinco dimensões.

Item	Rotação promax					Rotação varimax					Comuna- lidade
	F_1	F_2	F_3	F_4	F_5	F_1	F_2	F_3	F_4	F_5	
3	-0,356	<b>-0,514</b>	-0,020	0,176	0,077	<b>-0,434</b>	-0,391	0,003	0,040	0,101	0,353
5	0,072	-0,390	0,136	<b>1,007</b>	-0,327	0,208	0,030	0,198	<b>0,951</b>	-0,102	0,998
6	<b>-0,938</b>	-0,241	-0,020	-0,009	0,109	<b>-0,992</b>	-0,095	-0,022	-0,080	0,009	0,999
8	-0,071	-0,187	-0,052	<b>0,494</b>	-0,145	-0,004	0,038	-0,022	<b>0,465</b>	-0,054	0,221
10	-0,085	-0,178	-0,001	<b>-0,395</b>	-0,037	-0,115	-0,281	-0,042	<b>-0,390</b>	-0,063	0,251
12	-0,174	-0,299	-0,184	0,264	0,157	-0,251	-0,201	-0,143	0,141	0,167	0,171
19	-0,203	-0,220	0,060	-0,164	0,229	-0,330	-0,291	0,068	-0,263	0,176	0,298
21	-0,213	0,288	-0,017	-0,274	<b>0,815</b>	-0,510	0,021	0,040	-0,436	<b>0,584</b>	0,794
22	-0,233	-0,203	0,141	-0,021	0,026	-0,270	-0,173	0,140	-0,067	0,021	0,127
23	-0,164	-0,092	-0,005	-0,092	<b>0,392</b>	<b>-0,333</b>	-0,188	0,027	-0,219	0,305	0,288
24	-0,582	0,031	0,106	0,137	<b>0,689</b>	<b>-0,832</b>	0,005	0,184	-0,067	0,517	0,999
25	<b>-0,682</b>	0,360	0,041	0,466	0,065	<b>-0,617</b>	0,609	0,078	0,491	-0,014	0,999
27	0,018	-0,214	0,031	<b>-0,370</b>	-0,077	-0,003	-0,318	-0,010	<b>-0,364</b>	-0,075	0,239
28	0,068	0,121	-0,014	<b>-0,485</b>	0,245	-0,045	-0,121	-0,033	<b>-0,499</b>	0,143	0,287
29	0,058	-0,165	0,169	-0,286	-0,011	0,017	-0,269	0,143	-0,297	-0,009	0,182
30	-0,003	-0,286	-0,049	-0,204	<b>0,451</b>	-0,230	-0,455	-0,016	-0,380	<b>0,381</b>	0,550
32	-0,122	-0,202	0,071	-0,135	<b>0,320</b>	-0,283	<b>-0,301</b>	0,093	-0,261	0,262	0,316

Item	Rotação promax					Rotação varimax					Comuna- lidade
	F_1	F_2	F_3	F_4	F_5	F_1	F_2	F_3	F_4	F_5	
33	-0,008	<b>-1,148</b>	0,012	0,204	-0,178	-0,074	<b>-0,992</b>	0,021	0,011	0,008	0,990
35	-0,050	-0,227	-0,024	-0,139	0,075	-0,116	-0,275	-0,027	-0,196	0,067	0,132
37	-0,177	<b>-0,427</b>	0,083	0,044	0,140	-0,284	<b>-0,402</b>	0,103	-0,086	0,153	0,284
38	0,068	-0,041	0,078	-0,267	<b>0,486</b>	-0,150	-0,265	0,107	<b>-0,401</b>	0,387	0,415
40	-0,190	<b>-0,315</b>	-0,034	-0,079	0,014	-0,237	<b>-0,299</b>	-0,040	-0,139	0,017	0,167
43	0,206	<b>-0,323</b>	-0,120	0,179	0,217	0,087	<b>-0,337</b>	-0,074	0,035	0,256	0,194
45	-0,056	<b>-0,441</b>	0,005	-0,085	0,050	-0,137	<b>-0,456</b>	0,005	-0,182	0,077	0,266
46	-0,238	0,144	-0,010	<b>-0,412</b>	0,372	-0,387	-0,051	-0,013	<b>-0,462</b>	0,218	0,413
47	<b>-0,946</b>	-0,109	-0,029	-0,023	0,142	<b>-0,996</b>	0,021	-0,030	-0,076	0,019	0,999
48	-0,194	<b>-0,401</b>	0,052	-0,140	0,230	-0,342	<b>-0,459</b>	0,064	-0,277	0,201	0,449
52	0,017	-0,022	-0,057	-0,067	<b>0,566</b>	-0,213	-0,185	-0,001	-0,234	<b>0,463</b>	0,349
55	<b>-0,638</b>	0,216	0,060	0,621	-0,008	-0,556	0,533	0,105	<b>0,628</b>	-0,036	1,000
56	<b>-0,943</b>	-0,109	-0,022	-0,033	0,146	<b>-0,995</b>	0,016	-0,023	-0,086	0,021	0,999
57	-0,137	0,152	<b>-0,698</b>	-0,044	-0,033	-0,086	0,187	<b>-0,708</b>	0,005	-0,089	0,552
58	-0,013	0,083	<b>-0,792</b>	-0,547	-0,191	0,060	-0,031	<b>-0,864</b>	-0,429	-0,255	1,000
59	-0,130	<b>0,502</b>	0,004	-0,196	0,373	-0,222	<b>0,348</b>	0,020	-0,190	0,213	0,252
60	0,158	-0,020	<b>-0,993</b>	0,282	0,436	0,021	-0,027	<b>-0,915</b>	0,133	0,377	0,999
61	-0,143	0,352	-0,289	<b>-0,395</b>	0,017	-0,118	0,236	<b>-0,329</b>	-0,297	-0,097	0,275
64	-0,036	<b>1,163</b>	-0,014	-0,202	0,243	0,008	<b>0,998</b>	-0,017	-0,025	0,039	0,999
65	<b>-0,673</b>	0,643	0,013	0,137	0,169	-0,633	<b>0,744</b>	0,029	0,213	0,003	0,999
66	<b>-0,944</b>	-0,103	-0,025	-0,030	0,146	<b>-0,995</b>	0,023	-0,026	-0,083	0,021	0,999
69	-0,235	-0,248	-0,233	0,144	<b>-0,477</b>	-0,059	-0,022	-0,273	0,229	<b>-0,384</b>	0,278
70	0,076	0,229	-0,320	-0,131	<b>0,387</b>	-0,049	0,077	-0,290	-0,189	0,277	0,205
71	-0,348	-0,219	-0,330	-0,204	<b>-0,354</b>	-0,232	-0,120	<b>-0,391</b>	-0,123	-0,340	0,352
73	<b>-0,797</b>	0,466	0,195	0,067	-0,186	-0,646	<b>0,653</b>	0,164	0,217	-0,285	0,999
74	0,028	-0,089	<b>-0,869</b>	-0,138	-0,143	0,086	-0,074	<b>-0,895</b>	-0,098	-0,149	0,846
75	-0,009	-0,104	<b>-0,376</b>	0,057	-0,057	0,014	-0,053	<b>-0,376</b>	0,051	-0,043	0,149

A tabela da análise fatorial considerando cinco dimensões (Tabela 13) mostra que nesta configuração cinco itens (12, 19 22, 29 e 35) apresentaram carga fatorial inferior a 0,300 e, houve mais itens com sinais



contrários nas dimensões de maior carga fatorial, assim como na configuração tridimensional. Com exceção da dimensão 3, a qual repete praticamente os mesmos itens da configuração tetradimensional e tridimensional, as demais dimensões não apresentam um agrupamento conceitual definível, o que torna esta configuração pentadimensional pouco adequada empiricamente para o conjunto de dados em questão, independente do tipo de rotação utilizado.

Sendo assim, empiricamente o modelo mais apropriado para representar o construto qualidade em websites comerciais é o de quatro dimensões. Este modelo foi então considerado nas demais análises: teoria da resposta ao item uni e multidimensional e bi-fatorial.

#### 4.4 VERIFICAÇÃO DA ADEQUAÇÃO DO MODELO UNIDIMENSIONAL E UNIDIMENSIONAL MÚLTIPLO

Conforme descrito na seção 3.4.4, o primeiro modelo testado foi o logístico de dois parâmetros (equação 4). Para isto, foram utilizados os 44 itens e as 441 respostas. O software utilizado para fazer as estimações dos parâmetros do modelo foi o BILOG MG (MISLEVY; BOCK, 2003).

A Tabela 14 mostra as estimações dos parâmetros do modelo logístico de dois parâmetros unidimensional.

Observa-se que a estimação de 18 itens (em destaque na Tabela 14) é prejudicada, ou seja, ou apresentam problemas na magnitude das estimações e nos erros padrões ou apresentam parâmetro de discriminação inferior a 0,700, o que indica baixa informação do item no construto. Ao se comparar esta análise com a realizada na seção anterior, assumindo a análise fatorial com quatro dimensões (Tabela 11) verifica-se que a maioria dos itens eliminados pertencem à primeira e terceira dimensão que, segundo o discutido na seção anterior, agrupam itens relacionados a acessibilidade do sistema e controle por parte do usuário ou interação. Enquanto que os itens que estatisticamente estão alinhados com a abordagem unidimensional são itens na maioria associados a dimensão 2 e 4 da análise fatorial de quatro dimensões. Estas dimensões, conforme discutido associam apresentação da informação e navegação. Esta análise estatística e empírica reforça a inadequação do modelo unidimensional para o construto em análise.

Entretanto, partindo da análise da Tabela 11 e do discutido no parágrafo anterior realizou-se uma análise unidimensional múltipla considerando a existência de quatro construtos unidimensionais distintos. O

primeiro chamado de ModU1 considerando apenas os itens evidenciados na dimensão 1 da análise fatorial apresentada na tabela 10, um segundo, chamado de ModU2 considerando apenas os itens evidenciados na dimensão 2 da mesma análise, o mesmo foi feito com os itens evidenciados na terceira dimensão agrupados no modelo ModU3 e os itens do modelo ModU4 referem-se aos itens da dimensão 4.

Tabela 14 - Estimação dos parâmetros de dificuldade e discriminação assumindo o modelo unidimensional de dois parâmetros para os 44 itens.

Item	Parâmetros				Item	Parâmetros			
	Discrimi- nação	Erro Padrão	Dificul- dade	Erro Padrão		Discrimi- nação	Erro Padrão	Dificul- dade	Erro Padrão
3	0,917	0,310	-3,819	1,014	<b>43</b>	<b>0,491</b>	<b>0,400</b>	<b>7,714</b>	<b>5,915</b>
<b>5</b>	<b>0,085</b>	<b>0,000</b>	<b>-34,640</b>	<b>0,000</b>	45	0,962	0,159	-1,379	0,203
6	1,373	0,317	-2,694	0,455	46	1,474	0,215	-0,750	0,107
<b>8</b>	<b>0,001</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	47	1,163	0,345	-3,821	0,827
10	0,769	0,165	-2,333	0,442	48	1,795	0,264	-1,567	0,145
<b>12</b>	<b>0,388</b>	<b>0,143</b>	<b>-4,807</b>	<b>1,707</b>	52	1,094	0,168	-1,489	0,202
19	1,482	0,191	-0,979	0,112	55	0,783	0,331	-4,965	1,796
21	2,377	0,474	-1,981	0,178	56	2,181	0,615	-2,704	0,393
<b>22</b>	<b>0,625</b>	<b>0,232</b>	<b>-5,174</b>	<b>1,731</b>	<b>57</b>	<b>0,338</b>	<b>0,000</b>	<b>4,059</b>	<b>0,000</b>
23	1,397	0,262	-2,507	0,348	<b>58</b>	<b>0,366</b>	<b>0,000</b>	<b>3,847</b>	<b>0,000</b>
24	3,489	1,372	-2,366	0,228	<b>59</b>	<b>0,354</b>	<b>0,212</b>	<b>-7,256</b>	<b>4,181</b>
<b>25</b>	<b>0,646</b>	<b>0,203</b>	<b>-4,655</b>	<b>1,320</b>	<b>60</b>	<b>0,033</b>	<b>0,411</b>	<b>-106,467</b>	<b>1332,705</b>
27	0,761	0,131	0,697	0,175	<b>61</b>	<b>0,194</b>	<b>0,107</b>	<b>-1,403</b>	<b>0,907</b>
28	0,935	0,174	1,179	0,203	<b>64</b>	<b>0,330</b>	<b>0,548</b>	<b>-11,268</b>	<b>18,123</b>
29	0,808	0,150	2,176	0,373	<b>65</b>	<b>0,295</b>	<b>0,484</b>	<b>-12,891</b>	<b>20,643</b>
30	1,852	0,256	-0,714	0,087	66	1,112	0,421	-3,947	1,085
32	1,244	0,194	-1,273	0,171	<b>69</b>	<b>0,085</b>	<b>0,000</b>	<b>-10,520</b>	<b>0,000</b>
33	1,960	0,361	-2,319	0,252	<b>70</b>	<b>0,366</b>	<b>0,116</b>	<b>1,346</b>	<b>0,483</b>
35	0,799	0,158	0,439	0,146	<b>71</b>	<b>0,130</b>	<b>0,188</b>	<b>-13,665</b>	<b>19,541</b>
37	0,988	0,175	-2,005	0,294	<b>73</b>	<b>0,222</b>	<b>3,458</b>	<b>-21,259</b>	<b>327,392</b>
38	1,360	0,171	-0,452	0,095	<b>74</b>	<b>0,401</b>	<b>0,000</b>	<b>2,728</b>	<b>0,000</b>
40	0,831	0,143	0,174	0,127	<b>75</b>	<b>0,190</b>	<b>0,000</b>	<b>-2,269</b>	<b>0,000</b>

Os resultados da estimação dos parâmetros para os modelos unidimensionais ModU1, ModU2, ModU3 e ModU4 podem ser visualizado nas Tabelas 15, 16, 17 e 18, respectivamente.

Tabela 15 - Estimação dos parâmetros de dificuldade e discriminação assumindo o modelo unidimensional de dois parâmetros ModU1.

Item	Parâmetros			
	Discriminação	Erro Padrão	Dificuldade	Erro Padrão
57	1,77	0,26	1,16	0,13
58	6,68	3,79	0,80	0,20
60	1,93	1,01	-2,64	0,67
69	0,64	0,18	-1,52	0,39
71	1,00	0,29	-2,10	0,46
74	8,91	58,30	0,58	0,83
75	0,84	0,18	-0,59	0,16

Observa-se, na Tabela 15, que as estimatórias dos parâmetros tornaram-se mais estáveis que as estimatórias do modelo unidimensional geral. O item 74 apresentou um erro padrão mais alto que os demais itens para o parâmetro de discriminação, e, além disso, o parâmetro de discriminação de 8,91, na escala logística é considerado muito elevado, podendo indicar problemas de estimação. Com exceção do item 74 e do item 69 o qual apresentou parâmetro de discriminação inferior a 0,70, os demais itens apresentaram um comportamento estatístico melhor que o visualizado na Tabela 14 do modelo unidimensional geral.

Tabela 16 - Estimação dos parâmetros de dificuldade e discriminação assumindo o modelo unidimensional de dois parâmetros ModU2.

Item	Parâmetros			
	Discriminação	Erro Padrão	Dificuldade	Erro Padrão
3	1,25	0,43	-3,04	0,74
6	1,85	0,68	-2,30	0,51
12	0,76	0,22	-2,65	0,68
22	0,89	0,32	-3,85	1,16
23	1,32	0,34	-2,58	0,50
32	0,94	0,20	-1,59	0,30
33	2,58	0,69	-2,08	0,27
35	0,97	0,22	0,38	0,14
37	1,65	0,34	-1,47	0,20
40	0,68	0,17	0,20	0,16
43	1,25	0,60	3,54	1,18
45	1,07	0,23	-1,28	0,24
48	2,10	0,46	-1,47	0,17
56	2,07	0,64	-2,74	0,53
73	0,00	0,00	0,00	0,00

A estimação dos parâmetros do modelo ModU2, se comparada com a estimação dos parâmetros do modelo geral, mostra que com exceção do item 73, todos os demais itens que não apresentavam uma boa qualidade, com base no parâmetro de discriminação, passaram a apresentar um parâmetro de discriminação superior a 0,70. Na tabela 16 pode-se observar que o erro padrão, tanto do parâmetro de discriminação quanto do parâmetro de dificuldade manteve-se dentro de um padrão aceitável. Este resultado reforça a suposição de que a unidimensionalidade geral não é adequada ao conjunto dos 44 itens, mas pode ser adequada para subgrupos de itens como os apresentados pelo ModU2.

Tabela 17 - Estimação dos parâmetros de dificuldade e discriminação assumindo o modelo unidimensional de dois parâmetros ModU3.

Item	Parâmetros			
	Discriminação	Erro Padrão	Dificuldade	Erro Padrão
21	3,00	1,18	-1,85	0,25
24	9,10	505,90	-2,23	2,95
25	1,62	0,43	-2,39	0,45
46	1,39	0,32	-0,79	0,15
47	1,15	0,47	-3,86	1,36
52	1,10	0,26	-1,49	0,30
55	1,00	0,65	-4,09	2,44
59	0,96	0,37	-3,04	0,97
61	0,57	0,17	-0,51	0,25
64	1,31	0,93	-3,40	1,83
65	1,50	0,72	-3,22	1,08
70	0,86	0,21	0,65	0,18

O modelo ModU3, assim como o modelo ModU2 apresenta uma maior estabilidade das estimações dos parâmetros dos itens, Tabela 17, se comparado como os mesmos itens do modelo geral, tabela 14. Por exemplo, o item 59, no modelo unidimensional geral apresentou parâmetro de discriminação igual a 0,35 e na estrutura ModU3 apresentou parâmetro de discriminação igual a 0,96, o que representa um ganho de informação indicando maior aderência do item ao modelo ModU3. O mesmo acontece com os itens 25, 61, 64, 65 e 70. O único item deste subconjunto que apresentou instabilidade na estimação foi o item 24, o qual apresentou erro padrão igual a 505,90. Do ponto de vista empírico, pode-se perceber que a maioria dos itens do ModU3, conforme discutido na seção anterior, está ligada com a ideia de controle por parte do usuário e/ou interação deste com o sistema. O item 24, descrito como “existe um link de um único clique que conduz à homepage” parece apresentar uma ideia relacionada ao controle por parte do usuário, conforme discutido na seção anterior, entretanto num subconjunto reduzido e unidimensional, como o modelado pelo ModU3, esta característica pode não estar associada diretamente ao controle e sim ao conceito de navegação, o que justifica a instabilidade na estimação dos parâmetros. Ou ainda, esta instabilidade na estimação pode caracterizar a multidimensionalidade do item, esta suposição pode ser sustentada pela visualização da tabela 11, a qual traz as cargas fatoriais dos 44 itens

assumindo um modelo multidimensional de quatro dimensões. Nesta análise fatorial é possível verificar que apesar do item 24 possuir uma maior carga na dimensão 3 que foi estimada em 0,750, ele possui uma carga de igual sentido e de magnitude 0,520 na primeira dimensão e 0,320 na quarta dimensão.

Tabela 18 - Estimação dos parâmetros de dificuldade e discriminação assumindo o modelo unidimensional de dois parâmetros ModU4.

Item	Parâmetros			
	Discriminação	Erro Padrão	Dificuldade	Erro Padrão
5	0,01	2,09	-317,42	****
8	0,00	0,00	0,00	0,00
10	0,55	0,17	-3,53	1,01
19	0,86	0,19	-1,60	0,32
27	0,68	0,17	0,65	0,21
28	1,18	0,28	0,77	0,16
29	1,34	0,22	1,43	0,20
30	1,91	0,39	-0,97	0,15
32	1,10	0,23	-1,68	0,30
35	0,67	0,17	0,33	0,18
38	1,24	0,24	-0,73	0,16
66	0,74	0,41	-5,81	2,86

A estimação dos parâmetros dos itens no modelo ModU4, diferentemente dos outros 3 modelos tornou-se mais instável que a estimação dos mesmos itens considerando o modelo unidimensional geral. Por exemplo, os itens 10, 27 e 35 tiveram a magnitude dos parâmetros de discriminação reduzidos no ModU4. Particularmente o item 35, o qual foi utilizado no ModU2 e no ModU4, por apresentar carga fatorial igual tanto na dimensão dois quanto na dimensão quatro (ver Tabela 11), apresentou-se mais adequado ao subconjunto do modelo ModU2. Este resultado, do ponto de vista empírico faz bastante sentido, uma vez que o item 35 descrito como “em produtos em que existam mais de uma perspectiva, é possível visualizar todas as perspectivas” está conceitualmente mais alinhado a apresentação da informação, característica chave da dimensão dois da análise fatorial que deu origem ao subgrupo do ModU2 do que ao conceito de navegação e orientação do usuário, característica predominante no subgrupo de itens do modelo ModU4.

Em uma análise unidimensional é comum eliminar os itens com problemas de estimação e reestimar os demais itens para verificar se existe alguma alteração nos itens de boa qualidade, apesar de se assumir independência entre itens. Tal análise foi realizada nos quatro modelos e verificou-se poucas variações nas estimativas.

Outra alternativa de modelagem unidimensional, agora tentando agregar mais itens é a consolidação de dois modelos unidimensionais em um, tais como o modelo ModU1 e o ModU3 que teoricamente, agregam itens relacionados à acessibilidade e controle por parte do usuário/interação com o website, ou seja, características ligadas diretamente ao sistema. E os modelos ModU2 e ModU4 com itens associados a apresentação da informação e navegação/orientação do usuário, sendo portanto, itens ligados a parte visual e usual do sistema. Sendo assim, este agrupamento alternativo teve como base esta avaliação empírica do qual resultou o modelo ModU13, que agrupa os itens do ModU1 e do ModU3 e o modelo ModU24 que agrupa itens do ModU2 e ModU4.

Os resultados da estimação dos parâmetros para o modelo ModU24 pode ser visualizado na tabela 19, e os resultados para o modelo ModU13 pode ser visualizado na Tabela 20.

Tabela 19 - Estimação dos parâmetros de dificuldade e discriminação assumindo o modelo unidimensional de dois parâmetros ModU24.

Item	Parâmetros				Item	Parâmetros			
	Discrimi- nação	Erro Padrão	Dificul- dade	Erro Padrão		Discrimi- nação	Erro Padrão	Dificul- dade	Erro Padrão
3	1,049	0,337	-3,447	0,807	32	1,140	0,179	-1,364	0,196
6	1,539	0,377	-2,515	0,404	33	2,338	0,446	-2,148	0,214
10	0,837	0,178	-2,184	0,394	35	0,852	0,170	0,420	0,138
<b>12</b>	<b>0,489</b>	<b>0,158</b>	<b>-3,888</b>	<b>1,185</b>	37	1,263	0,215	-1,703	0,206
19	1,509	0,213	-0,975	0,117	40	0,966	0,162	0,155	0,113
21	2,115	0,415	-2,058	0,209	43	0,991	0,404	4,185	1,366
<b>22</b>	<b>0,656</b>	<b>0,248</b>	<b>-4,957</b>	<b>1,673</b>	45	1,052	0,178	-1,297	0,184
23	1,511	0,283	-2,391	0,323	47	1,385	0,384	-3,396	0,599
24	3,141	1,099	-2,390	0,260	48	1,815	0,295	-1,560	0,151
<b>25</b>	<b>0,686</b>	<b>0,230</b>	<b>-4,419</b>	<b>1,308</b>	52	1,005	0,175	-1,580	0,237
27	0,738	0,133	0,714	0,181	55	0,853	0,331	-4,620	1,495
30	1,645	0,245	-0,763	0,099	56	2,211	0,687	-2,674	0,409

Tabela 20 - Estimação dos parâmetros de dificuldade e discriminação assumindo o modelo unidimensional de dois parâmetros ModU13.

Parâmetros					Parâmetros				
Item	Discrimi- nação	Erro Padrão	Dificul- dade	Erro Padrão	Item	Discrimi- nação	Erro Padrão	Dificul- dade	Erro Padrão
<b>5</b>	<b>0,646</b>	<b>0,250</b>	<b>-7,580</b>	<b>3,025</b>	61	1,037	0,172	-0,319	0,113
<b>8</b>	<b>0,001</b>	<b>0,000</b>	<b>0,000</b>	<b>0,000</b>	<b>64</b>	<b>0,547</b>	<b>0,181</b>	<b>-6,959</b>	<b>2,321</b>
<b>28</b>	<b>0,416</b>	<b>0,105</b>	<b>2,248</b>	<b>0,580</b>	65	0,721	0,240	-5,562	1,736
<b>29</b>	<b>0,279</b>	<b>0,080</b>	<b>5,583</b>	<b>1,672</b>	<b>66</b>	<b>0,454</b>	<b>0,129</b>	<b>-8,507</b>	<b>2,589</b>
<b>38</b>	<b>0,213</b>	<b>0,057</b>	<b>-2,169</b>	<b>0,748</b>	<b>69</b>	<b>0,510</b>	<b>0,125</b>	<b>-1,856</b>	<b>0,453</b>
<b>46</b>	<b>0,367</b>	<b>0,087</b>	<b>-2,257</b>	<b>0,597</b>	<b>70</b>	<b>0,659</b>	<b>0,123</b>	<b>0,792</b>	<b>0,196</b>
57	1,979	0,300	1,085	0,111	71	1,027	0,225	-2,051	0,358
58	4,373	0,994	0,890	0,069	<b>73</b>	<b>0,643</b>	<b>0,244</b>	<b>-7,617</b>	<b>2,998</b>
<b>59</b>	<b>0,470</b>	<b>0,134</b>	<b>-5,549</b>	<b>1,570</b>	74	4,231	1,014	0,696	0,067
60	1,762	0,632	-2,724	0,526	75	0,746	0,145	-0,646	0,165

O modelo ModU24 considera apenas os 24 itens que na seção anterior tiveram maior carga fatorial nas dimensões 2 e 4, da estrutura tetradimensional. Neste caso, espera-se que o construto unidimensional meça o grau de navegação e apresentação da informação. Como pode ser visto na Tabela 19 este conjunto de itens mostrou um bom ajuste, considerando que todos os 24 itens puderam ser estimados e não houve nenhuma discrepância nos erros padrões, além disso, a maioria dos itens apresentou boa informação, ou seja, parâmetro de discriminação superior a 0,700. Isto indica que se estas características fossem tratadas separadamente poderia dar origem a uma estrutura unidimensional da forma como foi tratado em Tezza (2009) e Tezza et al., (2011).

Entretanto, o modelo ModU13 não se mostrou muito ajustado à estrutura unidimensional, como pode ser visto pela estimação dos parâmetros na Tabela 20. Verifica-se que o item 08 não pode ser estimado e que mais da metade dos itens apresentou parâmetro de discriminação inferior a 0,70. Isto indica que agrupar os itens que tratam de acessibilidade e controle por parte do usuário ou interação não gera um modelo unidimensional viável.

Este comportamento reforça a necessidade de se avaliar criteriosamente a dimensionalidade de um construto, pois a escolha de um modelo pode enviesar parte das conclusões. Além disso, esta análise



evidencia a necessidade de analisar em paralelo os resultados estatísticos e os resultados empíricos/práticos.

#### 4.5 - VERIFICAÇÃO DA ADEQUAÇÃO DO MODELO MULTIDIMENSIONAL COMPENSATÓRIO

Dadas as conclusões e verificações discutidas na seção anterior assumiu-se que o construto é composto de quatro dimensões. Sendo assim, trabalhou-se uma modelagem multidimensional de quatro dimensões utilizando a teoria da resposta ao item. Como foi discutido no capítulo 2, a TRI possui algumas similaridades com a análise fatorial, principalmente a modelagem multidimensional, entretanto, a TRI possui vantagens consideráveis frente a abordagem puramente fatorial principalmente pelo fato de tratar os itens de uma forma individual e não apenas em termos de agrupamento fatorial. Nesta linha é possível analisar além da qualidade do item, seu grau de dificuldade e com isso estimar scores aos respondentes. Segundo Bartolucci et al. (2012), nenhuma outra abordagem estatística permite estudar a dimensionalidade e o poder de discriminação de itens de uma maneira mais simples do que a TRI.

Para verificar a adequação do modelo de multidimensional compensatório, realizou-se primeiramente uma comparação estatística entre o modelo unidimensional da TRI na estrutura logística de dois parâmetros, e o modelo multidimensional compensatório. Esta primeira foi feita comparando-se os dois critérios de informação descritos da subseção 3.4.1, o critério de informação bayesiano (BIC) (SCHWARZ, 1978) e o critério de informação de Akaike (AIC) (AKAIKE, 1973). Esta comparação, apresentou um melhor ajuste ao modelo de quatro dimensões frente ao modelo de apenas uma dimensão. A diferença entre os valores do critério de Akaike (AIC) foi -359,871 e do critério Bayesiano (BIC) foi -1038,42, o que significa que em ambos os critérios houve ganho de informação ao se utilizar um modelo de quatro dimensões ao invés de um modelo de apenas uma dimensão. Este critério para avaliação do melhor modelo é utilizado por vários autores como, por exemplo, Rost (1997) e Kelderman (1997), Li e Lissitz (2012), estes últimos comparam o modelo unidimensional de dois parâmetros com o bi-fatorial.

A Tabela 21 mostra a estimação dos parâmetros dos 44 itens utilizando o modelo multidimensional da teoria da resposta ao item compensatório por meio do software flexMIRT<sup>TM</sup> (CAI, 2012).

Tabela 21 - Estimação dos parâmetros de discriminação para cada dimensão e parâmetro de discriminação multidimensional (MDISC), parâmetro de dificuldade escalar (d) o parâmetro de dificuldade multidimensional (MDIFF) e seus respectivos erros padrões (se) para os 44 itens, assumindo o modelo multidimensional de dois parâmetros compensatório.

Item	a 1	se	a 2	se	a 3	se	a 4	se	MDISC	d	se	MDIFF
03	0,30	1,03	<b>1,52</b>	3,50	0,07	39,49	-0,040	54,53	1,55	4,07	1,24	-2,62
05	-0,32	22,88	0,28	31,36	-0,40	34,43	<b>-0,880</b>	20,25	1,06	5,23	9,19	-4,95
06	1,09	1,70	<b>1,91</b>	15,07	0,30	54,56	-0,250	64,98	2,23	4,62	1,76	-2,07
08	-0,14	0,59	0,15	39,66	-0,41	35,75	<b>-1,210</b>	16,01	1,29	0,81	0,32	-0,63
10	-0,12	0,67	0,65	41,81	0,06	19,93	<b>1,140</b>	23,05	1,32	2,08	0,53	-1,58
12	-0,13	0,78	<b>1,03</b>	42,84	0,87	42,55	-0,680	23,18	1,52	2,49	0,58	-1,64
19	0,67	0,81	<b>0,90</b>	28,51	0,51	14,99	<b>0,920</b>	23,47	1,54	1,46	0,48	-0,95
21	1,43	1,46	0,89	38,57	<b>1,89</b>	15,48	0,970	30,04	2,71	5,12	2,41	-1,89
22	0,56	0,98	<b>0,82</b>	32,96	0,57	34,70	-0,590	19,83	1,29	3,75	1,02	-2,91
23	0,55	0,90	<b>1,08</b>	18,07	0,59	17,63	0,610	28,72	1,48	3,55	0,66	-2,40
24	11,44	177,69	5,62	257,53	<b>16,62</b>	283,21	7,090	291,11	22,11	46,63	719,38	-2,11
25	0,39	1,62	-0,11	99,18	<b>2,82</b>	24,08	-0,980	79,36	3,01	5,57	2,82	-1,85
27	0,06	0,56	0,62	40,23	-0,16	17,11	<b>1,010</b>	26,64	1,20	-0,60	0,34	0,50
28	0,15	0,72	0,04	52,30	0,42	41,15	<b>1,560</b>	10,25	1,62	-1,35	0,44	0,83
29	0,47	0,68	0,37	49,02	-0,36	23,97	<b>1,150</b>	22,52	1,34	-2,03	0,46	1,51
30	0,76	0,88	1,01	39,26	0,78	20,63	<b>1,290</b>	23,39	1,97	1,35	0,58	-0,69
32	0,62	0,65	<b>0,65</b>	18,39	<b>0,65</b>	10,41	0,610	14,00	1,27	1,58	0,42	-1,25
33	0,92	1,29	<b>2,29</b>	32,29	1,24	58,78	-0,010	61,27	2,76	5,55	2,87	-2,01
35	0,10	0,45	<b>0,64</b>	22,48	0,10	10,49	<b>0,640</b>	21,77	0,92	-0,35	0,30	0,38
37	0,62	0,77	<b>1,45</b>	11,39	-0,10	31,66	0,260	56,15	1,60	2,36	0,56	-1,47
38	0,95	0,70	0,47	33,23	0,70	21,06	<b>1,100</b>	10,75	1,68	0,67	0,48	-0,40
40	0,15	0,50	<b>0,90</b>	17,42	0,05	14,72	0,490	32,63	1,04	-0,15	0,28	0,14
43	0,01	0,86	<b>1,06</b>	9,27	-0,18	23,78	0,160	43,42	1,09	-4,20	1,18	3,86
45	0,28	0,66	<b>1,13</b>	25,77	0,05	17,52	0,710	40,90	1,36	1,49	0,41	-1,09
46	0,53	0,77	0,49	32,04	<b>1,32</b>	19,01	1,02	22,69	1,82	1,24	0,50	-0,68
47	0,60	3,15	1,07	48,33	<b>1,16</b>	41,31	-0,58	22,02	1,78	5,19	3,09	-2,91
48	0,90	1,04	<b>1,62</b>	20,43	0,79	28,94	0,67	44,95	2,12	3,06	0,80	-1,44
52	0,61	0,64	0,47	19,32	<b>0,95</b>	7,59	0,46	14,58	1,31	1,73	0,44	-1,32
55	0,45	0,92	0,23	41,62	<b>1,16</b>	16,83	-0,43	24,72	1,34	4,40	0,87	-3,29

Item	a 1	se	a 2	se	a 3	se	a 4	se	MDISC	d	se	MDIFF
56	1,91	4,34	<b>1,81</b>	18,87	0,93	41,42	0,33	49,98	2,81	6,88	16,99	-2,45
57	<b>-2,08</b>	0,72	-0,14	30,90	1,17	2,99	-0,03	35,93	2,39	-2,41	0,68	1,01
58	<b>-50,47</b>	398,55	8,56	309,51	6,92	181,90	10,11	217,53	52,64	-41,73	332,82	0,79
59	0,38	0,99	-1,19	27,54	<b>1,34</b>	46,79	0,69	76,93	1,96	3,80	1,45	-1,94
60	<b>-2,74</b>	4,63	-0,46	70,45	0,05	61,38	1,92	21,13	3,38	7,25	13,39	-2,15
61	-0,75	0,53	-0,40	21,09	<b>0,95</b>	25,62	0,62	38,81	1,42	0,38	0,28	-0,27
64	0,71	1,75	-1,92	38,71	<b>1,75</b>	56,92	0,33	115,42	2,71	6,39	3,54	-2,36
65	0,26	3,79	-2,53	56,83	<b>2,89</b>	84,56	0,98	160,92	3,97	8,67	20,84	-2,18
66	1,01	1,19	-0,14	27,72	0,74	29,10	<b>0,94</b>	24,34	1,57	4,87	2,89	-3,10
69	<b>-0,86</b>	0,56	0,32	12,34	-0,47	18,09	-0,40	23,44	1,11	1,11	0,37	-1,004
70	-0,35	0,44	-0,10	17,00	<b>0,82</b>	13,87	0,43	25,20	1,00	-0,57	0,27	0,573
71	<b>-1,02</b>	0,56	0,46	9,67	0,46	7,56	0,25	9,91	1,24	2,23	0,39	-1,81
73	0,82	11,08	<b>-3,37</b>	70,10	2,36	86,72	-0,22	185,21	4,20	10,71	37,53	-2,55
74	<b>-4,25</b>	1,69	0,60	25,59	1,23	10,37	0,62	19,79	4,51	-3,05	1,27	0,68
75	<b>-0,79</b>	0,37	0,29	15,64	0,33	13,21	-0,24	6,18	0,94	0,51	0,23	-0,54

A Tabela 21 mostra os parâmetros de discriminação de cada item em cada dimensão e o parâmetro de discriminação multidimensional (MDISC) de cada item, que é, segundo Reckase (1997), o análogo do parâmetro de discriminação unidimensional, conforme apresentado no capítulo dois como sendo a raiz quadrada da soma ao quadrado de cada item em cada dimensão. Desta forma quanto maior o parâmetro de discriminação multidimensional (MDISC), maior é o poder de discriminação multidimensional do item.

Além disso, a Tabela 21 mostra o valor do parâmetro de localização  $d$  de cada item que, segundo Reckase (1997), está relacionado à dificuldade do item. Porém, como discutido no capítulo dois, o valor deste parâmetro não pode ser interpretado da mesma maneira que o parâmetro de dificuldade do modelo unidimensional, pois no modelo multidimensional compensatório o parâmetro  $d$  é um escalar, portanto, apresentando apenas um único valor para cada item. O valor que é equivalente a interpretação unidimensional do parâmetro  $b$  é dada pela divisão de  $-d$  pelo parâmetro de discriminação multidimensional (MDISC), gerando assim, o parâmetro chamado dificuldade multidimensional (MDIFF) apresentada na última coluna da Tabela 21. O valor de MDIFF indica a distância da origem do

espaço ao ponto de inflexão na direção do vetor de discriminação. Sendo assim, quanto maior o valor do parâmetro MDIFF maior é a dificuldade do item.

O parâmetro de discriminação definido no capítulo 2, assim como a carga fatorial, pode ser arbitrário em cada dimensão, considerando que possa haver algum tipo de rotação, estes carregamentos podem ser comparados identificando itens semelhantes na mesma dimensão, da mesma forma como é feita a análise fatorial, utilizada na seção 4.4 (RECKASE, 1997). Sendo assim, espera-se que o parâmetro de discriminação de um determinado item seja maior na dimensão em que este tenha maior poder de discriminação, ou seja, na dimensão em que este possua mais informação. Sendo assim, avaliando-se os valores individuais dos parâmetros de discriminação em cada dimensão, verifica-se que a configuração das cargas foi a mesma apresentada na análise fatorial considerando quatro dimensões (Tabela 11). Entretanto, observa-se que os itens 24 e 58 apresentaram problemas de estimação tanto para o parâmetro de discriminação quanto para o parâmetro de dificuldade. Além disso, os erros padrões dos demais itens, principalmente nas dimensões 2, 3 e 4, apresentaram-se altos. Sendo assim, optou-se por eliminar os itens 24 e 58 e com isso verificar o comportamento das demais estimativas dos parâmetros e dos erros padrões. Além disso, nesta etapa, com o objetivo de reduzir os erros padrões melhorando assim as estimativas foi utilizado a distribuição beta ( $\alpha, \beta$ ) como *priori* para os parâmetros de discriminação, seguindo indicação de Chalmers (2012) e Houts e Cai (2012), neste caso com  $\alpha = 1,6$  e  $\beta = 0,1$ , detalhes sobre a distribuição beta podem ser encontrados em Rohatgi e Saleh (2001). Esta reestimação com 42 itens pode ser visualizada na Tabela 22.

Tabela 22 - Reestimação dos parâmetros de discriminação para cada dimensão e parâmetro de discriminação multidimensional (MDISC), parâmetro de dificuldade escalar (d) o parâmetro de dificuldade multidimensional (MDIFF) e seus respectivos erros padrões (se) para os 42 itens, assumindo o modelo multidimensional de dois parâmetros compensatório.

Item	a 1	se	a 2	se	a 3	se	a 4	se	MDISC	d	s.e.	MDIFF
3	-0,49	0,55	0,53	0,72	1,13	0,72	0,81	0,65	1,57	4,10	0,79	-2,62
5	-0,53	1,04	-0,30	1,27	0,01	0,77	0,31	0,91	0,68	4,92	1,09	-7,20
6	-0,71	0,31	1,17	0,66	0,84	0,54	1,32	0,89	2,08	4,44	1,11	-2,14
8	-0,93	0,29	-0,51	0,29	-0,28	0,33	-0,02	0,33	1,10	0,76	0,18	-0,69
10	0,48	0,37	0,39	0,32	0,95	0,43	0,36	0,40	1,19	1,99	0,30	-1,67
12	-0,29	0,09	0,00	0,60	-0,52	0,20	2,18	0,92	2,26	3,11	0,80	-1,38
19	0,49	0,39	1,12	0,38	1,19	0,40	0,46	0,45	1,77	1,55	0,26	-0,88

Item	a 1	se	a 2	se	a 3	se	a 4	se	MDISC	d	s.e.	MDFF
21	1,47	1,18	1,91	0,71	0,39	0,74	1,16	0,88	2,70	5,02	1,26	-1,86
22	-0,35	0,55	0,74	0,70	-0,39	0,64	1,05	0,65	1,39	3,88	0,86	-2,80
23	0,23	0,62	0,87	0,46	1,01	0,61	0,80	0,57	1,57	3,60	0,51	-2,29
25	0,89	0,60	0,39	0,54	-0,67	0,58	0,70	0,68	1,37	3,55	0,58	-2,59
27	0,38	0,41	0,52	0,39	1,46	0,46	-0,12	0,42	1,60	-0,67	0,23	0,42
28	0,99	0,34	0,68	0,34	0,51	0,34	0,16	0,35	1,31	-1,22	0,24	0,93
29	0,22	0,41	0,76	0,39	1,12	0,39	-0,25	0,44	1,39	-2,06	0,32	1,48
30	0,78	0,45	1,53	0,44	0,57	0,48	1,03	0,52	2,08	1,37	0,26	-0,66
32	0,46	0,37	0,96	0,35	0,38	0,36	0,65	0,37	1,30	1,58	0,24	-1,21
33	-0,14	0,87	1,26	0,72	0,88	0,87	2,33	1,26	2,79	5,53	1,58	-1,98
35	0,21	0,26	0,42	0,26	0,76	0,27	0,35	0,30	0,96	-0,36	0,17	0,38
37	-0,46	0,43	0,93	0,39	1,09	0,44	0,60	0,43	1,62	2,37	0,37	-1,46
38	0,91	0,45	1,72	0,46	0,06	0,46	0,67	0,49	2,06	0,74	0,23	-0,36
40	-0,02	0,27	0,50	0,27	0,83	0,29	0,46	0,33	1,07	-0,15	0,18	0,14
43	-0,25	0,62	0,17	0,55	0,74	0,72	0,78	0,65	1,12	-4,25	0,74	3,81
45	0,04	0,35	0,65	0,36	0,99	0,38	0,71	0,39	1,38	1,49	0,25	-1,08
46	1,19	0,41	0,96	0,38	0,43	0,43	0,81	0,42	1,78	1,21	0,24	-0,68
47	-0,25	0,70	0,64	0,90	0,28	0,14	1,24	1,11	1,45	4,72	1,02	-3,27
48	0,16	0,52	1,49	0,45	0,66	0,61	1,41	0,59	2,16	3,07	0,49	-1,42
52	0,75	0,40	1,01	0,39	-0,09	0,41	0,85	0,42	1,52	1,82	0,28	-1,20
55	0,25	0,17	0,55	0,51	0,24	0,68	0,19	0,61	0,68	3,79	0,64	-5,60
56	-0,16	0,84	1,58	0,91	1,14	1,08	1,26	1,25	2,33	5,98	1,72	-2,57
57	1,56	0,89	-2,48	1,00	0,28	0,82	1,30	0,87	3,22	-3,01	0,89	0,94
59	1,71	0,84	0,60	0,56	-0,56	0,61	-0,48	0,66	1,96	3,82	0,93	-1,95
60	1,39	1,13	-1,16	0,89	0,99	0,88	0,08	1,05	2,06	5,20	1,24	-2,52
61	1,44	0,41	-0,57	0,40	0,22	0,42	0,26	0,43	1,59	0,40	0,20	-0,25
64	1,85	1,12	0,72	0,78	-0,05	1,27	-1,26	1,03	2,35	5,91	1,60	-2,51
65	2,60	2,08	0,35	1,09	0,32	1,16	-1,24	1,49	2,92	6,97	3,54	-2,39
66	1,18	1,00	1,38	0,76	0,41	0,80	-0,20	0,96	1,87	5,23	1,63	-2,79
69	-0,49	0,36	-0,89	0,34	0,61	0,33	-0,13	0,40	1,19	1,14	0,24	-0,96
70	0,92	0,27	-0,11	0,29	0,13	0,33	0,29	0,33	0,98	-0,57	0,19	0,58

Item	a 1	se	a 2	se	a 3	se	a 4	se	MDISC	d	s.e.	MDIFF
71	0,50	0,39	-0,75	0,44	0,62	0,43	0,65	0,42	1,27	2,26	0,31	-1,78
73	1,88	1,81	0,74	1,42	-0,07	1,14	-1,70	1,88	2,64	7,87	4,07	-2,98
74	0,85	0,58	-1,65	0,55	0,55	0,58	0,82	0,57	2,10	-1,77	0,38	0,84
75	0,16	0,31	-0,77	0,29	0,07	0,33	0,72	0,35	1,07	0,54	0,17	-0,51

Observa-se na Tabela 22 que não houve nenhum problema na estimação dos parâmetros dos 42 itens. Entretanto, os itens 05 e 73 apresentaram erros padrões desproporcionais a magnitude dos parâmetros. Sendo assim, optou-se por retirar estes itens e reestimar os parâmetros dos demais itens. Os resultados desta reestimação podem ser visualizados na Tabela 23.

Tabela 23 - Reestimação dos parâmetros de discriminação para cada dimensão e parâmetro de discriminação multidimensional (MDISC), parâmetro de dificuldade escalar (d) o parâmetro de dificuldade multidimensional (MDIFF) e seus respectivos erros padrões (se) para os 40 itens, assumindo o modelo multidimensional de dois parâmetros compensatório.

Item	a 1	s.e.	a 2	s.e.	a 3	s.e.	a 4	s.e.	MDISC	d	s.e.	MDIFF
3	<b>1,43</b>	0,78	0,65	0,76	-0,34	0,64	0,66	0,74	1,74	4,23	0,82	-2,43
6	<b>2,29</b>	1,18	0,98	0,89	-0,03	0,70	0,39	0,93	2,52	4,88	1,39	-1,94
8	-0,02	0,35	0,29	0,40	<b>-0,99</b>	0,33	-0,70	0,39	1,25	0,79	0,19	-0,63
10	0,51	0,42	-0,15	0,43	0,39	0,42	<b>1,20</b>	0,46	1,37	2,09	0,36	-1,53
12	0,66	0,55	<b>1,47</b>	0,58	-0,06	0,57	-0,02	0,50	1,61	2,53	0,44	-1,57
19	<b>1,20</b>	0,39	0,14	0,45	0,75	0,38	0,92	0,40	1,69	1,50	0,25	-0,89
21	1,73	1,03	1,26	1,00	<b>2,35</b>	1,24	0,74	0,99	3,26	5,57	1,46	-1,71
22	<b>1,11</b>	0,79	0,94	0,71	0,33	0,63	-0,48	0,67	1,57	4,02	0,90	-2,57
23	<b>1,22</b>	0,50	0,42	0,62	0,44	0,59	0,91	0,63	1,64	3,61	0,52	-2,20
25	0,27	0,93	<b>2,39</b>	1,30	1,54	0,91	-0,61	0,99	2,92	5,34	1,50	-1,83
27	0,61	0,38	-0,39	0,41	0,24	0,40	<b>1,13</b>	0,40	1,36	-0,62	0,21	0,45
28	0,24	0,43	-0,31	0,43	<b>1,17</b>	0,40	1,01	0,45	1,59	-1,33	0,26	0,83
29	0,70	0,43	-0,71	0,46	0,39	0,47	<b>0,92</b>	0,43	1,41	-2,07	0,32	1,47
30	<b>1,40</b>	0,50	0,30	0,50	1,31	0,47	0,90	0,48	2,14	1,38	0,27	-0,65
32	<b>0,96</b>	0,36	0,38	0,41	0,86	0,40	0,47	0,39	1,42	1,62	0,24	-1,14
33	<b>2,50</b>	1,33	1,91	1,21	0,49	0,92	1,00	0,91	3,34	6,07	1,78	-1,82

Item	a 1	s.e.	a 2	s.e.	a 3	s.e.	a 4	s.e.	MDISC	d	s.e.	MDIFF
35	<b>1,60</b>	0,47	0,20	0,47	-0,13	0,42	0,61	0,49	1,73	2,43	0,37	-1,41
37	0,62	0,30	0,05	0,31	0,22	0,28	<b>0,77</b>	0,28	1,01	-0,36	0,17	0,36
38	1,16	0,44	0,04	0,49	<b>1,60</b>	0,47	0,39	0,48	2,01	0,73	0,24	-0,36
40	<b>0,82</b>	0,29	0,11	0,34	0,06	0,28	0,68	0,29	1,07	-0,15	0,18	0,14
43	<b>0,84</b>	0,53	0,42	0,69	-0,35	0,78	<b>0,75</b>	0,84	1,25	-4,36	0,91	3,48
45	0,68	0,75	<b>1,11</b>	0,76	0,60	0,70	-0,21	0,74	1,45	4,44	0,72	-3,06
46	0,70	0,40	0,59	0,46	<b>1,40</b>	0,45	0,83	0,41	1,87	1,23	0,25	-0,66
47	1,37	1,28	<b>1,51</b>	1,01	0,21	0,86	0,14	1,09	2,05	5,40	1,38	-2,63
48	<b>1,89</b>	0,58	0,73	0,59	0,81	0,54	0,72	0,62	2,30	3,12	0,50	-1,36
52	0,82	0,38	0,63	0,44	<b>1,14</b>	0,43	0,24	0,39	1,56	1,82	0,27	-1,17
55	<b>1,09</b>	0,41	0,14	0,42	0,14	0,38	0,99	0,40	1,49	1,52	0,27	-1,02
56	<b>2,57</b>	1,28	1,03	1,35	0,66	1,14	0,84	1,16	2,97	6,78	2,46	-2,28
57	-1,98	0,97	<b>2,04</b>	1,07	-0,26	0,81	1,63	0,98	3,29	-3,05	0,91	0,93
59	-0,55	0,59	0,27	0,73	<b>2,00</b>	0,91	-0,18	0,65	2,10	3,94	0,99	-1,88
60	-1,30	1,00	0,20	1,19	0,22	1,00	<b>2,03</b>	1,53	2,43	5,74	2,17	-2,36
61	-0,86	0,41	0,62	0,49	0,79	0,42	<b>0,90</b>	0,43	1,60	0,40	0,20	-0,25
64	-0,51	0,82	0,47	1,10	<b>2,08</b>	1,01	-0,34	1,29	2,22	5,59	1,34	-2,52
65	-1,00	1,05	0,86	1,46	<b>2,58</b>	1,92	0,19	1,28	2,90	6,71	2,89	-2,31
66	0,73	0,79	0,06	1,13	<b>1,56</b>	1,05	0,31	0,72	1,75	5,02	1,28	-2,87
69	-0,25	0,34	0,19	0,40	<b>-0,97</b>	0,31	0,30	0,34	1,06	1,10	0,24	-1,04
70	-0,29	0,30	0,48	0,35	<b>0,63</b>	0,31	0,51	0,32	0,99	-0,57	0,20	0,58
71	-0,31	0,45	0,79	0,47	-0,15	0,44	<b>1,01</b>	0,45	1,33	2,28	0,32	-1,72
74	-1,14	0,56	1,18	0,65	-0,36	0,54	<b>1,29</b>	0,65	2,12	-1,77	0,36	0,84
75	-0,35	0,31	<b>0,79</b>	0,35	-0,33	0,31	0,43	0,35	1,02	0,53	0,17	-0,52

A estrutura apresentada na Tabela 23 assemelha-se a apresentada na análise fatorial da tabela 10, a qual considera quatro dimensões. Entretanto, a configuração multidimensional da Tabela 23 apresenta algumas peculiaridades que a diferencia daquela. Primeiramente, quatro itens foram retirados por apresentarem problemas nas estimativas dos parâmetros. Em segundo, apesar de possuírem uma base estatística comum, análise fatorial de informação plena e a teoria de resposta ao item

multidimensional, a última trabalha com estimação de mais parâmetros o que justifica algumas particularidades no processo de estimação, as quais podem gerar problemas de convergência como as que ocorreram nos itens 05, 24, 58 e 73.

E por fim, a consequência gerada pelas reestimativas dos parâmetros é a possibilidade de reorganização dos itens nas dimensões e de alteração no significado base de alguma das dimensões, o que de fato ocorreu particularmente com a dimensão caracterizada na AF como acessibilidade e utilização do sistema e agora na reestimação da TRIM como acessibilidade e confiabilidade do sistema, representada pela dimensão dois.

Os valores em negrito na Tabela 23 destacam a dimensão na qual o item possui maior parâmetro de discriminação indicando assim em qual dimensão o item possui maior informação. Entretanto, como se trata de uma modelagem multidimensional compensatória, um item nem sempre é exclusivo de uma só dimensão, podendo assim, contribuir, para a estimação do traço latente do respondente (website) em várias dimensões.

De uma forma geral, a primeira dimensão, agrupa itens com características voltadas para condução/orientação do usuário no website também classificada, por alguns autores, como navegabilidade e usabilidade (GONZÁLEZ; PALACIOS, 2004; NUSAIR; KANDAMPULLY, 2008). Esta dimensão, contém itens com carregamento secundário principalmente na dimensão quatro, a qual é caracterizada por itens relacionados com apresentação da informação. Isto se deve ao fato de que, apesar de ser utilizada rotação ortogonal para apresentação dos resultados, existe algum tipo de relação entre as dimensões, e isto fica mais evidente nas dimensões um e quatro, particularmente porque estas características são tratadas em alguns estudos (TEZZA et al., 2011; YANG et al., 2004) como características altamente correlacionadas. Jun et al. (2004) classifica a organização da informação como facilidade de uso, que neste contexto pode-se entender como usabilidade. A dimensão três agrupa itens associados à interatividade.

No presente estudo, a correlação entre a dimensão um e a dimensão quatro foi de aproximadamente 0,4. O comportamento destas duas dimensões reflete no carregamento dos itens 03, 19, 23, 35, 37 e 43.

O item 03, descrito como “a homepage deixa claro o que o site faz, sem precisar usar a rolagem” possui claramente uma característica de organização da informação de modo a conduzir o usuário em sua decisão de permanecer no website ou desistir e ir para outro website, como pode ser visualizado na Tabela 23, o item 03 possui um parâmetro de discriminação de 1,43 na primeira dimensão e de 0,66 na dimensão quatro.



Esta estrutura compensatória permite que o aumento de qualidade em uma dimensão possa compensar a redução em outra. O item 19, descrito como “as palavras aparentemente clicáveis são de fato clicáveis” é nítido o conceito de navegação e orientação do usuário, conforme foi discutida na subseção 4.3, entretanto é inegável que este item assim como tanto outros possui reflexos em outros conceitos como, por exemplo, o de apresentação da informação, ou seja, a apresentação de uma palavra ou frase transmite uma informação de funcionalidade que orienta e informa o usuário na interface.

O item 23 possui um comportamento semelhante ao item 19, inclusive nos valores estimados dos parâmetros de discriminação nas dimensões um e quatro, 1,20 e 0,92 respectivamente para o item 19 e 1,22 e 0,91 respectivamente para o item 23. Descrito como “o logotipo da empresa está no canto superior esquerdo em todas as páginas do site” o item 23 caracteriza-se como orientativo porque informa o usuário de que o mesmo ainda está no website selecionado e, além disso, na grande maioria dos websites analisados, este logotipo possui além da função orientativa, uma função de navegação uma vez que este geralmente representa um *link* direto para a página inicial.

O item 35 (em produtos em que existam mais de uma perspectiva, é possível visualizar todas as perspectivas?) possui, conforme tabela 23, maior carga de discriminação na primeira dimensão ( $a=1,60$ ) e secundariamente na quarta dimensão ( $a=0,61$ ). Por ser uma característica que na maioria dos casos envolve algum tipo de funcionalidade de links que direcionam para imagens em três dimensões ou alguma sequência de imagens este item é primariamente classificado aqui como de orientação/navegação e por apresentar informações de produtos, secundariamente como pertencente também a dimensão quatro. O item 37 (os produtos possuem informações suficientes?) possui um comportamento semelhante, no entanto, neste item o maior carregamento é na dimensão quatro e secundariamente na dimensão um. Isto é justificável pelo fato de que neste caso, as informações em sua maioria são estáticas e dependem menos de questões navegacionais. Em contrapartida, o item 43 (o item possui multimídia para apresentação de produtos?) é uma característica, assim como o item 37, de maior carga navegacional. Da mesma forma o item 40 (o site apresenta produtos relacionados no final da página?).

Os itens 06 e 56 são itens predominantes da dimensão um, pois além de possuírem carga de discriminação maior nesta dimensão, são caracterizados como funções diretas de navegação na interface.

O item 30 (os resultados de busca permitem classificação por outros critérios além de custos?) possui maior carga na primeira dimensão,

e secundariamente, na terceira dimensão, a qual agrupa predominantemente itens relacionados à interatividade. Tal comportamento é justificável, uma vez que dar ao usuário mais opções de classificação envolve navegabilidade e possibilidade de interação com usuário com a interface.

A segunda dimensão evidenciou itens de acessibilidade, semelhante a dimensão um da análise fatorial da Tabela 11, e confiabilidade/segurança. Os itens característicos desta dimensão são os itens 12, 25, 45, 47, 57 e 75. O item 12 (existe informações de contato telefônico ou endereço) é uma característica, segundo Nielsen e Tahir (2002), importante para transmitir ao usuário confiabilidade e segurança, de que se ocorrer algum imprevisto ele tenha um endereço físico ou um telefone para recorrer. O item 25 (o site permite navegação em suas páginas em apenas uma janela, ou seja, não há abertura de novas janelas em meio a navegação) é um item ligado a confiabilidade e interação. Nielsen e Loranger (2006) afirmam que a maioria dos usuários não entende como manipular várias janelas e se concentra na que está mais a frente na tela. Se não puderem voltar, eles ficarão presos e confusos. Nos itens 45 e 47, a questão relacionada com segurança fica mais evidente uma vez que o primeiro item verifica se existe indicação de ambiente seguro no momento da efetivação da compra e o segundo verifica se o site possui outras formas de pagamento além do cartão de crédito. Os itens 57 e 75 conforme discutido na análise fatorial na seção 4.3 estão ligados diretamente com a acessibilidade do sistema.

A dimensão três, apresenta, semelhante a terceira dimensão da análise fatorial da seção 4.3, itens característicos de interatividade. Os itens 21 e 28 estão ligados diretamente ao mecanismo de busca do website, o qual representa uma ferramenta bastante comum para permitir ao usuário interagir com o sistema efetuando sua busca por produtos e/ou informações. O item 38, descrito como “existe a possibilidade de inserir *feedback* sobre os produtos por parte do consumidor” representa uma possibilidade de interação comumente utilizada em redes sociais e indicação/avaliação de produtos, além de representar uma via de condução/orientação do usuário e apresentação da informação representada pelo carregamento secundário na primeira dimensão. Os itens 46 (no preenchimento de formulários, o usuário consegue visualizar os próximos passos na interface?) e 52 (os dados obrigatórios são diferenciados dos dados opcionais de forma visualmente clara?) apesar de não serem itens puramente de interação estão envolvidos diretamente neste processo, pois o preenchimento de formulários é uma etapa interativa no processo de compra em um website. Estes carregam secundariamente na dimensão de quatro e um respectivamente, reforçando além da interatividade a apresentação da

informação e a condução/orientação do usuário. Os itens 64, 65 e 70 como discutidos na seção 4.3 refletem o controle e poder de interação do usuário para com o sistema possibilitando que aquele possa pausa ou interromper algum conteúdo interativo ou de atualização automática.

A quarta dimensão, como discutida anteriormente, possui correlação com a primeira dimensão, entretanto, existem itens característicos que a caracteriza como sendo apresentação da informação, como por exemplo, os itens 27, 37, 60 e 74. Os itens 27 (existe uma lista de perguntas frequentes – FAQs?), 37 (os produtos possuem informações suficientes) e 74 (etiquetas ou instruções são fornecidas quando o conteúdo exige a entrada de dados por parte do usuário) caracterizam-se por apresentação pura da informação. Os itens 10 (existe agrupamento de subcategorias?), 60 (a apresentação visual de texto e imagens de texto tem uma relação de contraste de no mínimo 4,5:1?) e 71 (Alterar a definição de um componente de interface de usuário não provoca, automaticamente, uma alteração de contexto, a menos que o usuário tenha sido avisado sobre essa situação antes de utilizar o componente?) versam sobre a forma de apresentar e organizar a informação. O item 61, como discutido na seção 4.3 trata da possibilidade de ampliar o texto em até 200% o que representa para o usuário comum uma possibilidade de interação e apresentação da informação (quarta dimensão) e para o usuário com algum tipo de deficiência representa acessibilidade (segunda dimensão).

Na grande maioria das aplicações da teoria da resposta ao item, seja ela unidimensional ou multidimensional, os parâmetros de dificuldade do modelo são positivos. Isto ocorre principalmente por questões relacionadas ao próprio construto, ou seja, assume-se que um valor positivo de parâmetro de discriminação resulta em uma monotonicidade crescente, ou seja, com o aumento da proficiência do indivíduo aumenta a probabilidade de resposta positiva ao item, este comportamento é bastante comum nas áreas educacional e psicológica, nas quais a TRI é mais difundida. Entretanto, é possível se ter valores do parâmetro de discriminação ( $a$ ) negativo, tal situação mantém o pressuposto da monotonicidade, entretanto ao invés de ser crescente passa a ser decrescente. Segundo Reckase (2009) em modelos multidimensionais da TRI não há necessidade de considerações quando o valor do parâmetro de discriminação for negativo mas, muito próximo de zero. No entanto, na presente abordagem, existem alguns valores de  $a$ s negativos que deve ser considerados, como por exemplo, o item 08 (o site possui menu em cascata?). De uma forma geral, verifica-se, na Tabela 23, que este item possui valores negativos para as dimensões um, três e quatro o que significa que quanto mais o website aumenta seu grau de qualidade nestas dimensões

menor é a probabilidade de o site possuir menu em cascata. A dimensão três, a qual caracteriza interatividade apresenta o maior valor ( $a=-0,99$ ), seguida da dimensão quatro ( $a=-0,70$ ). Possuir menu em cascata não é necessariamente uma unanimidade na literatura de qualidade em um website, em alguns casos, a utilização desta funcionalidade pode dificultar a interação de usuário com o website ou esconder informações úteis de usuário, o que neste caso justifica o comportamento decrescente deste item.

Outros itens também apresentam valores negativos para o parâmetro de discriminação em alguma dimensão. Como por exemplo, o item 29, o qual possui maior carga na dimensão quatro ( $a= 0,92$ ) e possui carga negativa na dimensão dois ( $a= -0,71$ ) indicando que a característica de dar sugestões às buscas realizadas pelos usuários tende a reduzir a confiabilidade do sistema, podendo ser encarada por alguns usuários como uma tentativa comercial pouco útil.

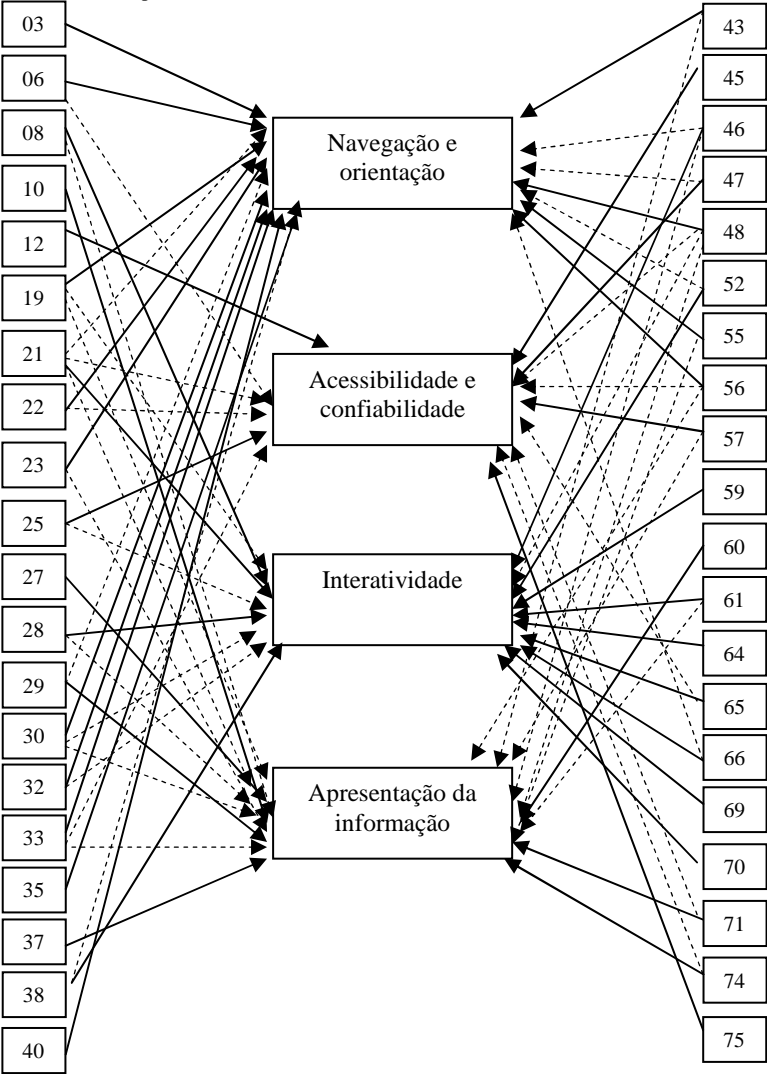
Dos itens que apresentaram alguma carga de discriminação negativa em alguma dimensão, a maioria pertence ao grupo de itens proposto pela WCAG 2.0, destinados a princípio a avaliarem automaticamente características do sistema associados com acessibilidade. Nota-se que destes itens (57, 59, 60, 61, 64, 65, 66, 69, 70, 71, 74 e 75), com exceção do item 66, todos apresentam parâmetro de discriminação negativo na primeira dimensão. Isto indica que para esta modelagem, tais características possuem uma redução na probabilidade de estarem presentes no website na medida em que o mesmo aumenta seu grau de proficiência na primeira dimensão. Isto pode ser explicado, em partes, pelo fato de que tais itens não interferem positivamente na condução/navegação de usuários sem deficiência visual ou cognitiva. Por exemplo, o item 57 (Todo o conteúdo não textual que é apresentado ao usuário tem uma alternativa em texto que serve um propósito equivalente?) só faz sentido se o usuário tiver algum tipo de deficiência visual ou em alguns casos específicos de segurança como, por exemplo, visualização de códigos especiais para dar continuidade a alguma operação. Sendo assim, é plausível que este item possua uma carga primária e positiva na segunda dimensão, a qual trata de acessibilidade e confiabilidade do sistema, mas, por consequência não tenha necessariamente uma carga positiva na primeira dimensão e nem na terceira, as quais tratam respectivamente de orientação/condução do usuário ou navegabilidade e interatividade. Tal comportamento se repete nos demais itens com menor intensidade dada a magnitude dos parâmetros quando comparado ao item 57.

Outra questão que pode surgir com relação a este grupo de itens é o fato de que por serem características avaliadas de forma automática alguns pontos não podem ser efetivamente avaliados, como por exemplo, nos itens

69 (Os cabeçalhos e as etiquetas descrevem o tópico ou a finalidade?), 74 (Etiquetas ou instruções são fornecidas quando o conteúdo exigir a entrada de dados por parte do usuário?) e 75 (No conteúdo implementado utilizando linguagens de marcação, os elementos dispõem de marcas de início e de fim completas, os elementos estão encaixados de acordo com as respectivas especificações, os elementos não contêm atributos duplicados, e todos os IDs são exclusivos, exceto quando as especificações permitem estas características?) é apenas possível verificar que o website possui apresentação da informação ou uma alternativa de acessibilidade, mas, não é possível verificar se tais etiquetas, instruções ou identificações são de fato informativas e não desorientativas ao usuário.

A Figura 14 mostra a estrutura multidimensional criada pelo modelo compensatório multidimensional da teoria da resposta ao item. Para construção deste diagrama, assumiu-se o carregamento em cada dimensão com base no sentido da carga de discriminação e na sua intensidade, assumindo aqui como sendo relevante cargas  $> 0,70$ . Segundo Hair et al., (2009), em uma análise fatorial, cargas fatoriais na faixa de  $\pm 0,30$  a  $\pm 0,40$  são consideradas como tendo o nível mínimo para interpretação de estrutura. Cargas de  $\pm 0,50$  ou maiores são tidas como praticamente significantes e maiores de  $\pm 0,70$  são consideradas indicativas de estrutura bem definida. Desta forma, optou-se por considerar como relevantes cargas maiores que  $0,70$ .

Figura 14 - Representação gráfica da relação de cada um dos 44 itens restantes com as quatro dimensões caracterizadas. A linha contínua representa a ligação de maior intensidade e as pontilhadas de menor.



A Figura 14 demonstra parte da complexidade envolvida em uma modelagem multidimensional compensatória, as linhas cheias, representam a interação de maior carga discriminatória e as linhas pontilhadas representam as interações secundárias com carga discriminatória no mesmo

sentido da principal e com intensidade maior que 0,70. Por exemplo, o item 47 (O site possui outras formas de pagamento além do cartão de crédito) possui maior discriminação na dimensão acessibilidade e confiabilidade com  $a = 1,57$  e interação secundárias na dimensão navegação e orientação com  $a = 1,37$ . Do ponto de vista prático representa que a característica possuir outras formas de pagamento além do cartão de crédito contribui do ponto de vista de qualidade em um website em termos de confiabilidade, ou seja, o website possibilita e permite ao usuário escolher a forma de pagamento que lhe pareça mais conveniente e segura, sendo, portanto uma característica comum às duas dimensões, como tantas outras características apresentada na figura 14.

O Quadro 6 mostra a distribuição dos itens quanto as dimensões caracterizadas, neste caso foram consideradas as maiores cargas.

Quadro 6 - Classificação dos itens conforme a dimensão, segundo o modelo multidimensional da teoria da resposta ao item.

<b>Dimensão</b>	<b>Itens</b>
Condução do usuário/navegabilidade	03, 06, 19, 22, 23, 30, 32, 33, 35, 40, 43, 48, 55, 56
Acessibilidade/confiabilidade do sistema	12, 25, 45, 47, 57, 75
Controle por parte do usuário/interatividade	08*, 21, 28, 38, 46, 52, 59, 64, 65, 66, 69*, 70
Apresentação da informação	10, 27, 29, 37, 60, 61, 71, 74

\*Itens com carregamento contrário a maioria dos itens da dimensão

Com base na estimação dos parâmetros de discriminação multidimensional, é possível classificar cada item conforme seu poder de discriminação ou seu poder informativo no construto, bem como com base no parâmetro de dificuldade classificar cada item conforme sua dificuldade. Esta classificação pode ser visualizada na Tabela 24.

Tabela 24 - Ranqueamento dos itens segundo os parâmetros de discriminação (MDISC) e dificuldade (MDIFF) de multidimensional.

Item	Rank Discriminação	Rank Dificuldade	MDISC	MDIFF	Item	Rank Discriminação	Rank Dificuldade	MDISC	MDIFF
3	18	35	1,74	-2,43	43	34	1	1,25	3,48
6	7	30	2,52	-1,94	45	28	40	1,45	-3,06
8	35	13	1,25	-0,63	46	16	15	1,87	-0,66
10	31	23	1,37	-1,53	47	14	38	2,05	-2,63
12	22	24	1,61	-1,57	48	9	21	2,30	-1,36
19	20	16	1,69	-0,89	52	26	20	1,56	-1,17
21	3	25	3,26	-1,71	55	27	17	1,49	-1,02
22	25	37	1,57	-2,57	56	4	32	2,97	-2,28
23	21	31	1,64	-2,20	57	2	3	3,29	0,93
25	5	28	2,92	-1,83	59	13	29	2,10	-1,88
27	32	7	1,36	0,45	60	8	34	2,43	-2,36
28	24	5	1,59	0,83	61	23	10	1,60	-0,25
29	30	2	1,41	1,47	64	10	36	2,22	-2,52
30	11	14	2,14	-0,65	65	6	33	2,90	-2,31
32	29	19	1,42	-1,14	66	17	39	1,75	-2,87
33	1	27	3,34	-1,82	69	37	18	1,06	-1,04
35	19	22	1,73	-1,41	70	40	6	0,99	0,58
37	39	8	1,01	0,36	71	33	26	1,33	-1,72
38	15	11	2,01	-0,36	74	12	4	2,12	0,84
40	36	9	1,07	0,14	75	38	12	1,02	-0,52

Na Tabela 24 verifica-se que os itens 43, 29, 57, 74 e 28 são os itens que exigem mais de um website, ou seja, são respectivamente os cinco itens mais difíceis do construto. O que, do ponto de vista prático faz sentido, primeiramente se avaliarmos o item 43 que é utilização de multimídia, por exemplo, vídeos, para apresentação de produtos, esta é uma característica pouco comum na maioria dos websites, encontrada apenas em websites mais maduros do ponto de vista de qualidade da apresentação da informação. O mesmo acontece com os itens 29 e 28 os quais tratam do aperfeiçoamento da ferramenta de busca.

Quanto ao rank de discriminação, na Tabela 24, verifica-se que os cinco itens mais discriminantes foram os itens 33, 57, 21, 56 e 25.



A Tabela 25 mostra a estimação da habilidade dos 4 primeiros websites analisados, na escala normal  $N(0 ; 1)$ , ou seja, média zero e variância um.

Tabela 25 - Estimação do grau de qualidade multidimensional dos 4 primeiros webistes da amostra com base no modelo multidimensional de dois parâmetros compensatório.

Website	Orientação do usuário/ navegabilidade	Acessibilidade/ confiabilidade do sistema	Controle por parte do usuário/ interatividade	Apresentação da informação
001	0,105	0,130	-0,864	0,677
002	0,367	0,524	-0,753	0,040
003	-0,887	-0,328	-0,557	0,525
004	0,782	0,324	-0,879	0,314

Verifica-se que o primeiro website possui um maior domínio dos itens relacionados à apresentação da informação, precisando assim, desenvolver melhor sua qualidade principalmente no controle por parte do usuário/interatividade, a qual apresentou domínio abaixo da média. No website número 002 observa-se que existe um bom domínio da acessibilidade/ confiabilidade do sistema, havendo, entretanto, necessidade de melhoramento no que se refere aos requisitos de controle por parte do usuário/interatividade.

Entretanto, em um modelo multidimensional, a análise pontual das estimações das proficiências dos respondentes, neste caso o grau de qualidade de um website pode remeter a uma avaliação simplista do modelo. Uma vez que a estrutura multidimensional agrega muito mais informação e complexidade que um modelo unidimensional. Sendo assim, cada dimensão pode ser avaliada em particular considerando cada item associado ao grau de qualidade estimado de cada website, como pode ser visualizado na Tabela 26. Esta tabela traz a probabilidade de acerto (resposta positiva) de cada item, com base nas estimativas do grau de qualidade dos quatro primeiros websites conforme Tabela 25.

Tabela 26 - Probabilidade de resposta positiva dos 40 itens para os 4 primeiros websites analisados com base nos seus respectivos graus de qualidade conforme estimado no modelo multidimensional compensatório de quatro dimensões.

Item	Website				Item	Website			
	001	002	003	004		001	002	003	004
3	0,995	0,995	0,964	0,998	43	0,032	0,028	0,009	0,046
6	0,996	0,998	0,940	0,999	45	0,982	0,992	0,954	0,991
8	0,770	0,839	0,710	0,820	46	0,675	0,685	0,518	0,731
10	0,931	0,876	0,891	0,922	47	0,997	0,999	0,975	0,999
12	0,944	0,973	0,815	0,973	48	0,961	0,974	0,756	0,987
19	0,835	0,815	0,612	0,892	52	0,762	0,832	0,593	0,851
21	0,988	0,994	0,937	0,996	55	0,900	0,873	0,721	0,931
22	0,975	0,991	0,908	0,991	56	0,999	1,000	0,986	1,000
23	0,983	0,982	0,932	0,990	57	0,159	0,080	0,276	0,039
25	0,981	0,996	0,958	0,992	59	0,888	0,914	0,958	0,856
27	0,488	0,324	0,360	0,469	60	0,999	0,995	1,000	0,995
28	0,158	0,096	0,173	0,124	61	0,579	0,463	0,729	0,381
29	0,141	0,080	0,100	0,141	64	0,973	0,983	0,990	0,968
30	0,740	0,750	0,446	0,846	65	0,990	0,992	0,997	0,982
32	0,793	0,824	0,601	0,868	66	0,981	0,985	0,975	0,987
33	0,999	1,000	0,970	1,000	69	0,895	0,864	0,876	0,871
35	0,959	0,962	0,792	0,983	70	0,324	0,293	0,365	0,262
37	0,511	0,440	0,344	0,547	71	0,959	0,939	0,948	0,939
38	0,435	0,497	0,269	0,591	74	0,365	0,223	0,433	0,174
40	0,589	0,548	0,357	0,665	75	0,764	0,747	0,729	0,719

Do ponto de vista prático, a análise mostrada na Tabela 26 demonstra, com base no grau de qualidade do website estimado pelo modelo multidimensional compensatório de quatro dimensões, os itens que são teoricamente dominados pelo website. Considerando o conceito de itens âncoras definidos por Beaton e Allen, (1992) e Andrade et al. (2000) no qual um dos critérios para caracterizar um item em determinado grau da escala é a probabilidade maior do que 0,650 naquele ponto da escala, verifica-se que o website 001 domina a maioria todos os itens, exceto os itens 27, 28, 29, 37, 38, 40, 43, 57, 61, 70 e 74, os quais apresentam probabilidade de resposta positiva inferior a 0,650. Observa-se que apesar

de o website 001 apresentar maior grau de qualidade na dimensão 4, ele tem, assim como os outros 3 websites mostrados na tabela 26, probabilidade abaixo de 0,650 ( $p=0,511$ ) de responder positivamente ao item 37 o qual trata, segundo a análise do modelo, de uma característica associada a apresentação da informação. Isto se dá particularmente por se tratar de um item com parâmetro de dificuldade acima da média (0,360). Tal análise pode ser seguida aos demais itens.

A análise de cada website associado com a probabilidade de resposta positiva a cada item amplia a análise prática da modelagem multidimensional, pois particulariza cada item e demonstra onde precisa ser melhorado pontualmente. Já a análise geral do grau de qualidade apresentado na Tabela 25 dá um indicativo relativo a dimensão ou ao conceito no qual o website precisa melhorar ou explorar mais.

#### 4.6 - VERIFICAÇÃO DA ADEQUAÇÃO DO MODELO BI-FATORIAL

A diferença básica entre o modelo bi-fatorial e o modelo multidimensional compensatório da TRI é a de que o primeiro apresenta um fator geral no qual todos os itens carregam em outras dimensões específicas. Mais detalhes sobre as diferenças entre estes modelos foram discutidos no capítulo 2.

O modelo bi-fatorial mais difundido na literatura é o modelo de natureza confirmatória. Casos específicos de análise exploratória Jenrich e Bentler (2011), vêm sendo desenvolvidos, entretanto, sua aplicação prática ainda é limitada. Desta forma, a abordagem adotada neste trabalho foi a confirmatória. Para isso, foram tomadas como base as definições do número de dimensões e dos agrupamentos de itens definidos na seção 3.4.3 (acessibilidade, apresentação da informação, interatividade, navegabilidade) assumindo-se como dimensão geral a qualidade de websites comerciais.

A Tabela 27 mostra as estimações dos parâmetros no modelo bi-fatorial, assumindo a estrutura confirmatória com base nas dimensões encontradas na seção 4.3.

Tabela 27 - Estimação dos parâmetros de discriminação (a) para cada dimensão e na dimensão geral e parâmetro de dificuldade escalar (d) assumindo o modelo bi-fatorial.

Item	Geral	$a_1$	$a_2$	$a_3$	$a_4$	$d$
03	0,531	0	0,374	0	0	2,116
05	-0,483	0	0	0	-0,178	-0,178
06	0,755	0	0,676	0	0	2,342

Item	Geral	$a_1$	$a_2$	$a_3$	$a_4$	$d$
08	-0,447	0	0	0	-0,839	0,538
10	0,522	0	0	0	0,747	1,293
12	0,106	0	0,831	0	0	1,418
19	0,948	0	0	0	0,021	0,832
21	1,804	0	0	1,104	0	3,364
22	0,233	0	0,730	0	0	2,2
23	0,790	0	0,333	0	0	2,03
24	6,115	0	0	4,418	0	12,924
25	0,385	0	0	0,997	0	2,252
27	0,482	0	0	0	0,151	-0,332
28	0,632	0	0	0	0,567	-0,768
29	0,505	0	0	0	0,218	-1,074
30	1,286	0	0	0	0,428	0,822
32	0,805	0	0	0,068	0	0,916
33	1,146	0	1,127	0	0	3,182
35	0,456	0	0,412	0	0	-0,233
37	0,597	0	0,648	0	0	1,317
38	0,953	0	0	0	0,151	0,353
40	0,519	0	0,076	0	0	-0,103
43	0,277	0	2,097	0	0	-4,691
45	0,627	0	0,274	0	0	0,793
46	0,999	0	0	0,584	0	0,69
47	0,672	0	0	0,402	0	2,591
48	1,089	0	0,722	0	0	1,79
52	0,709	0	0	0,253	0	0,956
55	0,388	0	0	0,741	0	2,519
56	1,310	0	0,321	0	0	3,426
57	-0,203	-1,064	0	0	0	-1,209
58	-0,513	-3,581	0	0	0	-2,966
59	0,243	0	0	0,900	0	1,925
60	-0,045	-1,175	0	0	0	2,986

Item	Geral	$a_1$	$a_2$	$a_3$	$a_4$	$d$
61	0,102	0	0	0,495	0	0,161
64	0,120	0	0	1,796	0	3,903
65	0,102	0	0	2,676	0	5,447
66	0,690	0	0	0	-0,155	2,572
69	-0,324	-0,387	0	0	0	0,616
70	0,228	0	0	0,411	0	-0,337
71	0,051	-0,635	0	0	0	1,253
73	-0,103	0	-0,694	0	0	3,168
74	-0,253	-2,751	0	0	0	-1,829
75	-0,091	-0,509	0	0	0	0,296

Na estrutura bi-fatorial é possível verificar o quanto os itens estão associados ao fator geral, neste caso representado como a qualidade de um website. Analisando-se os valores das cargas associadas ao fator geral, verifica-se que a maioria dos itens identificados no modelo anterior como característicos de requisitos do sistema, tais como acessibilidade, apresentaram baixa carga no fator geral, e, além disso, algumas como por exemplo, os itens 57, 58, 60, 69, 74 e 75 apresentaram carga negativa, isto matematicamente reflete o carregamento negativo visualizado tanto nos fatores secundários quando no modelo multidimensional apresentado na seção 4.5. No entanto, se verificarmos os carregamentos secundários nestes itens, existe uma uniformidade da intensidade e o sentido dos parâmetros, podendo assim indicar que características inerentes do sistema possuem uma orientação diferente das características associadas à organização da informação ou navegação direta, podendo, portanto, representar características ortogonais ou não compensatórias e que não podem ser tratadas como pertencentes a um fator geral. Particularmente, o fato do modelo bi-fatorial assumir ortogonalidade entre as dimensões secundárias e destas em relação ao fator geral, limita a adequação deste modelo a construto que possua claramente um fator geral ortogonal às demais subdimensões, o que não é o caso do construto em questão. Desta forma, verifica-se que qualidade em websites comerciais não é uma característica que possa ser representada por uma dimensão geral ortogonal às demais subdimensões, ao menos não para o construto desenvolvido neste trabalho. Sendo assim, a comparação destes três modelos sugere que qualidade de websites comerciais é uma característica não unidimensional, desdobrável em quatro dimensões compensatórias.

O ajuste do modelo bi-fatorial com relação ao modelo compensatório da TRIM, ambos assumindo-se quatro dimensões, foi avaliado com base nos critérios de informação BIC e AIC, RMSEA e M2 e pode ser verificado na Tabela 28.

Tabela 28 - Comparação entre os modelos unidimensional, multidimensional de dois parâmetros compensatório e do modelo bi-fatorial com base nos critérios de informação AIC e BIC, RMSEA e M2.

Modelo	Log de máxima verossimilhança	AIC	BIC	RMSEA	M2
Modelo Unidimensional	13826	14002,7	14362,6	0,07	2844,39
Modelo TRIM	12616	13016,1	13833,9	0,04	1405,22
Modelo Bi-fatorial	13200	13464,1	14003,8	0,05	1937,95

Verifica-se na Tabela 28 que, tanto no critério AIC quanto no BIC, o modelo multidimensional compensatório de quatro dimensões (TRIM), apresentou menor valor do que o modelo bi-fatorial e o modelo unidimensional, o que indica que aquele modelo é mais adequado aos dados que os modelos bi-fatorial e unidimensional. Esta verificação é confirmada por meio da raiz do erro médio quadrático de aproximação (RMSEA) o qual indica um menor erro para o modelo TRIM. A estatística M2, disponível no software flexMirt e discutida por Joe e Maydeu-Olivares (2010) e Liu e Maydeu-Olivares (2012), assemelha-se à estatística qui-quadrado e vem sendo amplamente utilizada em verificação de modelo da teoria da resposta ao item, sendo que, quanto menor for o seu valor, mais ajustado é o modelo em comparação aos demais. Desta forma, por este critério, o modelo TRIM é o mais adequado dos três.

#### 4.7 SÍNTESE DO CAPÍTULO

As três primeiras seções deste capítulo discutiram os resultados da análise de dimensionalidade do conjunto de itens e a qualidade destes em relação ao contruto geral. Nestas seções verificou-se a complexidade envolvida em uma análise de dimensionalidade. Considerando-se que

dependendo da técnica estatística empregada para tal análise os resultados podem divergir em termos de número de dimensões. Na presente análise, o método de informação restrita sugeriu a existência de 9 dimensões. Já o método de informação plena sugeriu a existência de 3 dimensões enquanto a técnica da análise paralela indicou a existência de mais do que 3 dimensões.

Sendo assim, o equacionamento destas divergências de resultados foi feito com base na análise empírica das dimensões e do conceito dos itens associados a cada uma. Resumidamente a definição do número de dimensões do construto foi feito com base na interpretação teórica de cada dimensão em relação aos itens associados a ela, resultando portanto, em uma estrutura tetradimensional.

Com base nesta definição, nas três últimas seções, partiu-se para a análise dos modelos unidimensional e multidimensional da teoria da resposta ao item e o modelo bi-fatorial. Na análise do modelo unidimensional pôde-se verificar que não se trata de um modelo adequado para o construto pelo fato de existirem dimensões diferentes no conjunto o que infringe os pressupostos básicos do modelo, incluindo a independência local. Entretanto, uma análise unidimensional múltipla, a qual subdivide o conjunto geral de itens em subconjuntos unidimensionais com base nas dimensões definidas anteriormente, mostrou-se mais adequado. A abordagem unidimensional possui algumas vantagens e desvantagens. Como vantagem de uma abordagem unidimensional ou unidimensional múltipla é a facilidade de análise e de representação da escala resultante. Entretanto, uma desvantagem da suposição de unidimensionalidade em um construto multidimensional é o fato de que o resultado será uma combinação linear das dimensões, o que pode não representar a realidade. Na utilização de uma estrutura unidimensional múltipla existe o inconveniente de se gerar  $k$  escalas diferentes e teoricamente não comparáveis em termos de parâmetros com média e desvio padrão. Isto dificulta a análise conjunta de todos os itens.

Já o modelo multidimensional, além de ter se mostrado estatisticamente mais adequado do que o unidimensional para este trabalho, conforme análise da Tabela 28, o mesmo possui mais possibilidades de análise conjunta, pois considera o construto como um todo particularizado em dimensões. Esta análise conjunta permite, conforme discutido na seção 2.3.3 gerar uma série de medidas relativas aos itens, tais como discriminação multidimensional, dificuldade multidimensional e localização multidimensional. Além de mensurar na mesma escala as proficiências (grau de qualidade) dos websites em cada dimensão particular.

Sob este ponto de vista, a TRIM sobressai-se à TRI unidimensional devido a sua possibilidade de análise conjunta de cada item e de cada respondente em cada dimensão, e como consequência é possível identificar a probabilidade de possuir determinada característica em cada website, com base na qualidade estimada dos parâmetros dos itens, como pode ser verificado na Tabela 26. O modelo multidimensional da TRI traz consigo grandes possibilidades de análise e interpretações, entretanto, estas vantagens resultam num aumento da complexidade particularmente por se estar trabalhando analiticamente no espaço vetorial e, portanto com múltiplas associações geométricas difíceis de serem visualizadas e interpretadas nas formas analíticas tradicionais.

O modelo bi-fatorial possui algumas peculiaridades, discutidas na seção 2.3.3, as quais podem limitar sua aplicação. A principal delas é a suposição de ortogonalidade das dimensões secundárias entre si e em relação ao fator geral. Na presente análise pôde-se perceber (Tabela 27) que a estimação do fator geral é prejudicada particularmente nos últimos itens, o que talvez tenha relação com o fato da primeira dimensão secundária (acessibilidade) possuir cargas fatoriais em sentido contrário das demais e do fator geral. Além disso, itens associados à dimensão secundária número quatro (controle por parte do usuário/interatividade) também tiveram estimação do fator geral prejudicada. Esta verificação sugere que os dados não possuem um fator geral ortogonal às demais dimensões secundárias. Do ponto de vista estatístico a adequação deste modelo em comparação com os modelos multidimensional e unidimensional pode ser visto na Tabela 26, na qual verifica-se que o modelo bi-fatorial possui um melhor ajuste em relação ao unidimensional porém, um pior ajuste em relação ao modelo multidimensional (TRIM).

De uma forma geral verificou-se neste capítulo que a análise de dimensionalidade e verificação da qualidade dos itens é fundamental para definir os possíveis modelos que possam se ajustar à realidade dos dados. Nem sempre o modelo unidimensional é o mais indicado, assim como nem sempre o modelo multidimensional ou bi-fatorial é a melhor escolha. Depende principalmente da natureza dos itens e das características dos respondentes. Merecendo, portanto, uma análise aprofundada de cada elemento.



## 5. CONCLUSÃO

### 5.1 CONCLUSÕES

A sistemática proposta neste trabalho visou levantar questões teóricas e práticas sobre qualidade em webistes comerciais, com a finalidade de construir um modelo multidimensional utilizando a teoria da resposta ao item. Desta forma, o primeiro passo se deu pela definição de qualidade em websites comerciais com base em um levantamento sistemático da literatura. Tal levantamento mostrou um campo aberto e sem definições diretas, tratando muitas vezes de estudos isolados e dificilmente comparáveis em termos de instrumento de medida. Para dar sequência ao desenvolvimento da sistemática, foi necessário particularizar o conceito de qualidade do websites à qualidade das funcionalidades deste, uma vez que todo o desenvolvimento foi baseado no sistema e suas peculiaridades e não na interação e percepção do usuário ao sistema. Este conceito possibilitou, com base em estudos anteriores, o desenvolvimento de um conjunto de 75 itens objetivos para avaliar qualidade de um website comercial. Com base neste desenvolvimento, realizou-se uma aplicação direta destes itens a uma amostra de 441 websites com o objetivo de fazer um levantamento de dados que retratasse a realidade do mercado de e-commerce para a partir daí desenvolver um modelo capaz de mensurar e orientar os websites no sentido de aumentar sua qualidade e consequentemente seus ganhos financeiros. De uma forma geral, como a literatura sobre qualidade de websites trata esta característica como multidimensional, no entanto sem definir uma quantidade única de dimensões, tão pouco a caracterização destas dimensões, realizou-se um estudo da dimensionalidade. Este estudo utilizou em paralelo análise estatística e análise empírica resultando em uma estrutura composta de quatro dimensões. Apesar desta caracterização encontrar respaldo em estudos anteriores assumiu-se que o desenvolvimento de um modelo multidimensional não deve finalizar com a definição das dimensões e sim com o tratamento individual de cada dimensão e de cada item com o objetivo de mensurar a característica fim, que no caso é a qualidade de websites comerciais.

Para isso, deu-se continuidade ao tratamento dos dados considerando os modelos multidimensionais da teoria da resposta ao item. Por meio desta análise, verificou-se a inadequação da utilização de um modelo unidimensional geral ou de um modelo multi-unidimensional, o qual utiliza  $k$  modelos unidimensionais para tentar representar um construto

geral. Verificou-se, entretanto a adequação do modelo multidimensional de dois parâmetros compensatório bem como a qualidade dos 75 itens, que ao final resultaram em um conjunto de 40 itens e a representação destes quanto a sua dificuldade. A estimação destes parâmetros permitiu a estimação de um score representando o grau de qualidade de cada website em cada uma das quatro dimensões. Como análise adicional verificou-se, com base no modelo multidimensional compensatório o ajuste dos dados ao modelo confirmatório bi-fatorial. Esta análise mostrou que estatisticamente o modelo multidimensional agrega mais informação ao construto, quando comparado ao modelo bi-fatorial e ao modelo unidimensional. Desta forma, verifica-se que o modelo bi-fatorial não representa mais informação ao construto necessitando de uma abordagem possivelmente diferente a dada neste trabalho.

Sendo assim, o trabalho desenvolveu uma modelagem multidimensional utilizando a teoria da resposta ao item para mensurar qualidade em websites comerciais. Tal modelagem utilizou como base o modelo logístico de dois parâmetros multidimensional com quatro dimensões e um conjunto de 40 itens. As quatro dimensões mostraram-se associadas a conceitos de navegabilidade ou condução/orientação do usuário, acessibilidade e confiabilidade do sistema, interatividade e apresentação da informação. Estas dimensões encontradas no presente trabalho remetem as dimensões mais citadas na literatura e relacionadas diretamente com a definição de qualidade em websites adotada no levantamento bibliográfico do capítulo 2. A qual define qualidade em websites como um conjunto de características técnicas e não técnicas de um sistema web, que permitam ao usuário realizar seus objetivos em um website de forma acessível, eficiente e agradável. Como características técnicas entende-se usabilidade/navegabilidade, apresentação da informação, acessibilidade e interatividade do sistema. Como características não-técnicas entende-se design, estética, apelo visual/comercial, confiabilidade, hedonismo, empatia.

Do ponto de vista de implicações teóricas ao campo de qualidade em websites, o presente trabalho desenvolveu uma sistemática para avaliar requisitos de qualidade, discutiu questões relativas à dimensionalidade, e sugeriu um modelo multidimensional capaz de avaliar individualmente cada item e cada website em suas respectivas dimensões.

Do ponto de vista prático, o trabalho abriu várias possibilidades de mensuração da qualidade em websites comerciais com a finalidade de identificar possíveis melhorias à interface e possivelmente o desenvolvimento de ferramentas automatizadas para diagnósticos úteis para melhoria de projetos de websites.

## 5.2 RECOMENDAÇÕES

No presente trabalho a elaboração dos itens, e identificação *a priori* das dimensões foi feita com base na literatura e não envolveu testes com usuário ou entrevistas para identificar satisfação e/ou preferências destes. Sendo, portanto uma avaliação objetiva do sistema web. Desta forma, a aplicação da metodologia apresentada neste trabalho em testes com usuários ou questionários para identificar preferências ou satisfação dos mesmos é uma oportunidade ainda inexplorada e pode representar um valioso avanço para a área.

Além disso, o processo de compras em websites pode ser encarado como uma cadeia de processo muito mais ampla do que simplesmente o sistema web, como foi abordado neste trabalho. Alguns autores como Li et al. (2002), Parasuraman et al. (2005) entre outros encaram o processo de compras online desde o primeiro contato com o websites, até o processo de pós venda, passando pela negociação e pelo processo logístico de transporte físico do material. Desta forma, sugere-se a ampliação da abordagem realizada neste trabalho ao processo de compras online como um todo, contribuindo assim não só para a área de web design ou de sistema web mas também para a área de logística e marketing.

Um dos limites deste trabalho foi a utilização apenas de websites de língua portuguesa. Este limite poderia ser retirado de uma abordagem futura, possibilitando assim, avaliar influências culturais e linguísticas na análise, abrindo campo para a possível aplicação da técnica de DIF (*Differential Item Functioning*).

Outro limite do trabalho foi a utilização somente de websites comerciais, entretanto, existe uma gama de outros websites com características e objetivos diferentes, com algum tipo de interesse comercial indireto como, por exemplo, as publicidades em websites de notícias ou entretenimento. Tais websites possuem conceitos de qualidade diferentes do abordado neste trabalho e apontam como uma alternativa para aplicação da metodologia apresentada neste trabalho.

As recomendações acima indicam outras possíveis abordagens tomando como base a abordagem aqui apresentada, entretanto, é possível também desenvolver outras abordagens ou até mesmo melhorar esta abordagem. Por exemplo, a presente abordagem traz um conjunto de itens com 40 itens, como se trata de avaliação de um sistema e a coleta de dados não depende exclusivamente de usuários, este conjunto de itens poderia ser ampliado, com o objetivo de identificar possivelmente mais dimensões ou

conceituar melhor as dimensões encontradas. Além disso, um aumento no número de itens e consequentemente na qualidade destes pode auxiliar no desenvolvimento de uma alternativa automática para fazer a coleta dos dados e consequentemente aplicar um Teste Adaptativo Computadorizado (TAI) para fazer a avaliação e o diagnóstico da qualidade de um website comercial. Abordagem esta que só é viabilizada pela metodologia da teoria da resposta ao item. Ainda nesta linha de continuidade do trabalho, é possível explorar outros modelos alternativos como, por exemplo, modelo hierárquico *testlets* proposto por Wainer et al., (2007) no qual é possível estimar uma dimensão geral hierárquica e correlacionada às demais subdimensões.

## REFERÊNCIAS

- ACKERMAN, T. A. A didactic explanation of item bias, item impact, and item validity from a multidimensional perspective. **Journal of Educational Measurement**, v.29, p.67-91, 1992.
- ACKERMAN, T. A. Using multidimensional item response theory to understand what items and tests are measuring. **Applied Measurement in Education**, v.7, p.255-278, 1994.
- ADAMS, R. J., WILSON, M., WANG, W.C. The Multidimensional Random Coefficients Multinomial Logit Model, **Applied Psychological Measurement**, v.21, n.1, p.1-23, 1997
- AGARWAL R., VENKATESH V., Assessing a Firm's Web Presence: A Heuristic Evaluation Procedure for the Measurement of Usability, **Information Systems Research**, v.13 n.2, p.168-186, 2002.
- ALADWANI, A.M. PALVIA, P.C. Developing and validating na instrument for measuring user-perceived web quality. **Information & Management**, v.39, p. 467–476, 2002.
- ALBA, J., LYNCH, J., WEITZ, B., JANISZEWSKI, C., LUTZ, R. SAWYER, A., WOOD S.. Interactive home shopping: consumer, retailer, and manufacturer incentives to participate in electronic marketplaces. **Journal of Marketing**, v.61, n.3, p. 38–53, 1997.
- AL-KHALIFA, H.S. The accessibility of Saudi Arabia government Web sites: an exploratory study, **Universal Access in the Information Society**, v.10, n.4, 2010.
- ALLEN, D. D., WILSON, M. Introducing multidimensional item response modeling in health behavior and health education research, **Health Education Research**, v.21 (Suppl. 1), p.73-84, 2006.
- ANDRADE, D. F.; TAVARES, H. R. e VALLE, R. C. **Teoria de Resposta ao Item: conceitos e aplicações**. ABE — Associação Brasileira de Estatística, 4º SINAPE, 2000.
- ANDRICH, D. Application of a psychometric rating model to ordered categories which are scored with successive integers. **Applied Psychological Measurement**, v.2, p.581–594, 1978.
- ANSLEY, T. N., FORSYTH, R. A. An examination of the characteristics of unidimensional IRT parameter estimates derived from two-dimensional data. **Applied Psychological Measurement**, v.9, n.1, p.37-48, 1985.

BABBIE, E., **The Basics of Social Research**. Wadsworth Publishing, Belmont, CA, 2005.

BABCOCK, B. G. E., **Estimating a Noncompensatory IRT Model Using a Modified Metropolis Algorithm**, PhD dissertation, University of Minnesota, 2009.

BAGOZZI, R. P., GOPINATH, M., NYER, P. U. The role of emotions in marketing. **Journal of the Academy of Marketing Science**, 27, 184–206, 1999.

BAKER, D.L., Advancing E-Government performance in the United States through enhanced usability benchmarks. **Government Information Quarterly**, v.26, n.1, p.82–88, 2009.

BAKER, F. B. **The Basics of Item Response Theory**. 2 ed. USA: ERIC Clearinghouse on Assessment and Evaluation, 2001. Disponível em: <http://edres.org/irt/> Acessado em 12 agosto 2010

BAKER, F.B., KIM, S.-H., **Item Response Theory: Parameter Estimation Techniques**, second ed. Marcel Dekker, New York, 2004.

BANGOR, A. KORTUM, P. T.. MILLER, J. T. An empirical evaluation of the system usability scale. **International Journal of Human-Computer Interaction**, v.24, n.6 p.574-594, 2008.

BARNES, S.J., R.T. VIDGEN. WebQual: An Exploration of Web Site Quality, **Proceedings of the Eighth European Conference on Information Systems**, v.1 p.298-305, Vienna, July 2000.

BARNES, S.J., VIDGEN, R.T. An integrative approach to the assessment of E-commerce quality, **Journal of Electronic Commerce Research** v.3, n.3, p. 114–127, 2002.

BARTOLUCCI, F. A class of multidimensional IRT models for testing unidimensionality and clustering items. **Psychometrika**, v. 72, p.141–157, 2007.

BASTIEN, J. M. C., SCAPIN, D. L. **Ergonomic criteria for the evaluation of human-computer interfaces**. Rapport technique de l'INRIA. 1993. Disponível em : <http://www.inria.fr/rrrt/rt-0156.html>.

BAUER, H.H., FALK T., HAMMERSCHMIDT, M. ETransQual: a transaction process-based approach for capturing service quality in online shopping. **Journal of Business Research**, v.59, p.866–875, 2006.

BAUER, H.H., HAMMERSCHMIDT, M., FALK, T., Measuring the quality of e-banking portals, **International Journal of Bank Marketing**, v.23 n.2, p.153-75, 2005.

BAUER, C., SCHARL A. Quantitive evaluation of Web site content and structure. **Internet Research: Electronic Networking Applications and Policy**, v.10, n.1, p. 31-43, 2000.

BEATON, A. E.; ALLEN, N. L. Interpreting scales through scale anchoring. **Journal of Educational Statistics**, v. 17, p. 191-204, 1992.

BIRNBAUM, A. Some Latent Trait Models and Their Use in Inferring an Examinee's Ability. In F. M. Lord and M. R. Novick. *Statistical Theories of Mental Test Scores*, MA: Addison-Wesley. goodness of fit test for the Rasch model. **Psychometrika**, v.38, p. 123-140, 1968.

BOCK, R. D. A Brief History of Item Theory Response. **Educational Measurement Issues and Practice**, v.16, n.4, p.21-33, 1997.

BOCK, R. D. Estimating item parameters and latent ability when responses are scored in two or more nominal categories. **Psychometrika**, v.37, p. 29-51, 1972.

BOCK, R., D., AITKIN, M. Maginal maximum likelihood estimation of item parameters: Application of an EM algorithm. **Psychometrika**, v.46, p.443-445, 1981.

BOCK, R. D., GIBBONS, R., MURAKI, E. Full information item factor analysis. **Applied Psychological Measurement**, v.12, p.261-280, 1988.

BOLLEN, K.A. **Structural Equations with Latent Variables**, New York: John Wiley & Sons, Inc. 1989.

BORTOLOTTI, S. L. V., **Resistência à mudança organizacional**: medida de avaliação por meio da teoria da resposta ao item. Florianópolis, SC, 2010 291p. Tese (doutorado) - Universidade Federal de Santa Catarina, Centro Tecnológico. Programa de Pós-Graduação em Engenharia de Produção.

BRADLOW, E.T., WAINER, H., WANG, X. A Bayesian random effects model for testlets. **Psychometrika**, v.64, p.153-168, 1999.

BRAZ, C., SEFFAH A. M'RAIHI D. Designing a Trade-Off Between Usability and Security: A Metrics Based-Model, **Lecture Notes in Computer Science**, v.4663, p.114-126, 2007.

BROSSMAN, B. G. **Observed score and true score equating procedures for multidimensional item response theory**. PhD dissertation, University of Iowa, 2010. Disponível em <http://ir.uiowa.edu/cgi/viewcontent.cgi?article=1654&context=etd>, Acessado em 18 março de 2011.

BUSCH, C., MARET, P. S. D., FLYNN, T., KELLUM, R., LE, S., MEYERS, B.,. **Content Analysis**. : Writing@CSU. Colorado State University Department of English. 2005. Disponível em : <http://writing.colostate.edu/guides/research/content/> Acessado em 02/04/2011.

BÜYÜKÖZKAN, G., RUAN, D., FEYZIOGLU, O. Evaluating e-learning web site quality in a fuzzy environment. **International Journal of Intelligent Systems**, v.22, n.5, p.567-586, 2007.

CAI, L. **flexMIRT™ version 1.86**: A numerical engine for multilevel item factor analysis and test scoring. [Computer software]. Seattle, WA: Vector Psychometric Group, 2012

CAI, S., JUN, M. Internet users' perceptions of online service quality: a comparison of online buyers and information searchers, **Managing Service Quality**, v.13, n.6, p. 504–519, 2003.

CAMILLI, G. A conceptual analysis of differential item functioning in terms of a multidimensional item response model. **Applied Psychological Measurement**, v.16, p.129-147, 1992.

CHALMERS, R.P., mirt: A Multidimensional Item Response Theory Package for the R Environment, **Journal of Statistical Software**, v.48, n.6, p.1-29, 2012.

CHANG, H.H., CHEN, S.W., Consumer perception of interface quality, security, and loyalty in electronic commerce. **Information & Management**, v.46, p.411–417, 2009

CHANG, M.K., CHEUNG, W., LAI, VINCENT.S. Literature Derived Reference Models for the Adoption of Online Shopping, **Information & Management**, v.42, n.4, p. 543-559, 2005.

CHARNES, A.; COOPER, W. W. & RHODES, E. Measuring the efficiency of decision making units, **European Journal of Operations Research**, v.2, p. 429-444, 1978.

CHAU, P.Y.K., HU, P.J., LEE, B.L.P. AU, A.K.K. Examining customers' trust in online vendors and their dropout decisions: an empirical study, **Electronic Commerce Research and Applications**, v. 6 n. 2, p. 171-82, 2007.

CHEN, F. F., WEST, S. G., SOUSA, K. H. A comparison of bifactor and second-order models of quality of life. **Multivariate Behavioral Research**, v.41, p.189–224, 2006.

CHERNYSHENKO, O. S., STARK, S., CHAN, K.-Y., DRASGOW, F., WILLIAMS, B. Fitting item response theory models to two personality inventories: Issues and insights. **Multivariate Behavioral Research**, v.36, p.563–562, 2001.

CHIOU, W.-C., LIN, C.-C., PERNG C., A strategic framework for website evaluation based on a review of the literature from 1995–2006, **Information & Management**, v.47, p. 282–290, 2010.

CHOI, J., LEE, K.H. **Differences in risk perception between apparel and non apparel online purchase: a cross-cultural study**”, paper presented at International Textiles and Apparel Association Conference, New York, NY, November, 2002.

CHRISTOFFERSSON, A. Factor analysis of dichotomized variables. **Psychometrika**, v.40, p.5-32, 1975.



- COBER, R., BROW, D., LEVY, P., Form, content, and function: an evaluative methodology for corporate employment web sites, **Human Resource Management**, v.43 n. 2/3, p. 201-18, 2004.
- COLLIER J.E., BIENSTOCK, C.C. Measuring service quality in e-retailing, **Journal of Service Research**, v.8, n.3, p.260–275, 2006.
- COMPUTING, T.R. **Terra Research and Computing**. The Galileo Computer Program [Computer software], Terra Research and Computing, Birmingham, 1994.
- COLESCA, S.E. An Assessment of the Quality of the Romanian Urban Web Sites, **Informatica Economică**, v.42, n.2, 2007.
- CRISTOBAL, E., FLAVIAN C., GUINALIU, M. Perceived e-service quality (PeSQ): measurement validation and effects on consumer satisfaction and web site loyalty, **Managing Service Quality**, v.17 n.3, p. 317–340, 2007.
- CRONIN JR., J. J., TAYLOR, S. A. Measuring service quality: a reexamination and extension. **Journal of Marketing**, v.56, n.3, p.55–68, 1992.
- CYBIS, W. **Ergonomia e Usabilidade** : conhecimentos, métodos e aplicações / Walter Cybis, Adriana Holtz Betiol, Richard Faust. São Paulo : Novatec Editora, 2007.
- DAY, A. A model for monitoring Web site effectiveness. **Internet Research: Electronic Networking Applications and Policy**, v.7, n.2 p.1-9, 1997.
- DAVEY, T., HIRSCH, T. M. **Concurrent and consecutive estimates of examinee ability profiles**. Paper presented at the Annual Meeting of the Psychometric Society, New Brunswick, NJ, 1991.
- de AYALA, R.J., **The Theory and Practice of Item Response Theory**. Guilford Press, New York, 2009.
- DeLONE, W.H., McLEAN E.R. The DeLone and McLean model of information systems success: a ten-year update. **Journal Management Information Systems**, v.19, n.4, p.9-30, 2003.
- DENG, Liqiong, POOLE, Marshall Scott. Aesthetic design of e-commerce web pages – Webpage Complexity, Order and preference. **Electronic Commerce Research and Applications**, v.11, n.4, p. 420–440, 2012.
- DESMARTEAU, K., Online apparel sales see double-digit growth, **Apparel Magazine**, v. 45 n.11, p.30-33, 2004.
- DeVELLIS, R.F. Classical Test Theory, **Medical Care**, v.44, n.11, Supl. 3, p.S50-S59, 2006
- DICKENSON, T. S. **Comparison of various ability estimates to the composite ability best measured by the total test score**. PhD Dissertation, University of South Carolina, 2005.

DING, D. X., HU, P.J.H., SHENG, O. R. L. e-SELFQUAL: A scale for measuring online self-service quality. **Journal of Business Research**, v.64, n.1, p.508–515, 2011.

DIVGI, D.R. Calculation of the tetrachoric correlation coefficient. **Psychometrika**, 44, n. 2, 169-172, 1979.

DOERFEL, M.L. BARNETT, G.A. A semantic network analysis of the international communication association, **Human Communication Research**, v.25 n.4, p.389– 603, 1999.

DOERFEL, M.L. BARNETT, G.A. A semantic network analysis of the international communication association, **Human Communication Research** v.25, n.4, p.389– 603, 1999.

DOWNING, S. M. Item response theory: Applications of modern test theory in medical Education, **Medical Education**, v.37 p.739–745, 2003.

DOWNES, P. K. Creating a practice website. **British Dental Journal**, v.202, n.10, p.597–604, 2007.

E-BIT. **Web Shoppers**. 26<sup>a</sup> ed. 2012. Disponível em: <http://www.webshoppers.com.br/webshoppers/WebShoppers26.pdf> Acessado em 17 setembro 2012.

EMBRETSON, S. E. A multidimensional latent trait model for measuring learning and change. **Psychometrika**, v.56, p.495-515, 1991.

EMBRETSON, S.; REISE, S. P. **Item Response Theory for Psychologists**. New Jersey: Lawrence Erlbaum Associates, Inc. Publishers, 2000.

ETHIER, J., HADAYA, P., TALBOT, J., CADIEAUX, J. Interface design and emotions experienced on B2C Web sites: Empirical testing of a research model. **Computers in Human Behavior**, v. 24, n.2, p. 2771-2791, 2008.

FANG, X., HOLSAPPLE, C.W. An empirical study of web site navigation structures' impacts on web site usability, **Decision Support Systems**, v.43,n.2, 2007, p. 476-491.

FASSNACHT, M., KOESE, I. Quality of electronic services: conceptualizing and testing a hierarchical model, **Journal of Service Research**, v.9, n.1, p.19–37, 2006.

FERGUSON, G. A. Item selection by the constant process, **Psychometrika**, v.7 n.1, 1942.

FERREIRA, S.B.L, NUNES, R.R. **e-Usabilidade**, Rio de Janeiro: LTC, 2008.

FERREIRA, O.G.L., MACIEL, S.C., SILVA, A.O.,NOVA SÁ, R.C. MOREIRA, M.A.S.P., Significados atribuídos ao envelhecimento: idoso, velho e idoso ativo, **Psico-USF**, v. 15, n. 3, p. 357-364, 2010.

FLAVIÁN, C., GUINALÍU, M. Consumer trust, perceived security and privacy policy: Three basic elements of loyalty to a web site, **Industrial Management & Data Systems**, v.106 n.5, p.601 – 620, 2006.

FLAVIÁN, C., GUINALÍU, M., GURREA, R., The influence of familiarity and usability on loyalty to online journalistic services: the role of user experience, **Journal of Retailing and Consumer Services**, v.13 p.363-75, 2006.

FLECK, M. P.A., BOURDEL, M. C. Método de simulação e escolha de fatores na análise dos principais componentes, **Revista Saúde Pública**, v.32, n.3, p267-72, 1998.

FORD, J. K., MACCALLUM, R. C., TAIT, M.. The application of exploratory factor analysis in applied psychology: a critical review and analysis. **Personnel Psychology**, v.39, p.292-314, 1986.

FOX, J.-P. **Bayesian Item Response Modeling: Theory and Applications**. New York: Springer, 2010.

FRAGOSO, T.M. **Modelos Multidimensionais da Teoria da Resposta ao Item**, Dissertação de mestrado, ICMC-USP, 2010.

FRANCIS J.E., WHITE, L. **PIRQUAL**: a scale for measuring customer expectations and perceptions of quality in Internet retailing. In: K. Evans and L. Scheer, Editors, Marketing Educators' Conference: Marketing Theory and Applications vol. 13 American Marketing Association, Chicago, IL, p. 263–270, 2002.

FRANKLIN, S.B., GIBSON, D.J., ROBERTSON, P.A., POHLMANN, J.T. FRALISH, J.S.. Parallel analysis: a method for determining significant principal components. **Journal of Vegetation Science**. v.6, n.1 p.99-106, 1995.

FRASER, C., McDONALD, R. P.. NOHARM: Least squares item factor analysis. **Multivariate Behavioral Research**, v.23, 267–269, 1988.

FREUND, J.E.; SIMON, G.A. **Estatística básica**: economia, administração e contabilidade. 9 ed. Porto Alegre: Bookman, 2000.

FROHLICH, D.. Beauty as a design prize. **Human-Computer Interaction**, Special Issue on Beauty Goodness and Usability, v.19, n.4, p.359 – 366, 2004.

GALLETTA, D. F.; HENRY, R.; MCCOY, S.; POLAK, P. Web Site Delays: How Tolerant are Users?, **Journal of the Association for Information Systems**. v.5, n.1, PP.1-28, 2004

GEFEN, D. Consumer loyalty in e-commerce. **Journal Association Information Systems** v.3, p.27–51, 2002.

GEISLER, G., ZINKHAN, G., WATSON, R. Web home page complexity and communication effectiveness. **Journal of Association for Information Systems**, v.2, p.1–46, 2001.

GIBBONS, R. D., HEDEKER, D. R. Full-information item bi-factor analysis. **Psychometrika**, v.57, p.423–436, 1992.

GIBBONS, R. D., BOCK, R. D., HEDEKER, D., WEISS, D. J., SEGAWA, E., BHAUMIK, D. K., KUPFER, D. J., FRANK, E., GROCHOCINSKI, V. J., STOVER, A. Full-information item bi-factor analysis of graded response data. **Applied Psychological Measurement**, v.31, p.4–19, 2007.

GIBBONS, R. D., IMMEKUS, J. C., BOCK, R. D. **The added value of multidimensional IRT models**. Multidimensional and hierarchical modeling monograph 1. Chicago: Center for Health Statistics, University of Illinois, 2007.

GIBBONS, R.D. RUSH AND, A.J. IMMEKUS, J.C. On the psychometric validity of the domains of the PDSQ: an illustration of the bi-factor item response theory model, **Journal of Psychiatric Research** v.43, p.401–410, 2009.

GIL, A. C. **Como elaborar projetos de pesquisa**. 4º ed São Paulo: Atlas, 2002.

GLAS, C. A. W. **A Rasch model with a multivariate distribution of ability**. In M. Wilson (Ed.), *Objective Measurement: Theory into practice* (pp. 236-258). Norwood NJ: Ablex, 1992.

GOUNARIS, S., DIMITRIADIS, S. Assessing service quality on the web: evidence from business-to-consumer portals, **Journal of Services Marketing**, v.17 n.4/5, p.529–548, 2003.

GOLDBERG, C. B., ALLEN, D. G.. Black and white and read all over: Race differences in reactions to recruitment Web sites. **Human Resource Management**, v. 47 n.2, p.217 – 236, 2008.

GONZÁLEZ, F.J.M. PALACIOS, T.M.B. Quantitative evaluation of commercial websites: an empirical study of Spanish firms, **International Journal of Information Management**, v.24, p.313–328, 2004.

GOUNARIS, S. DIMITRIADIS, S. Assessing service quality on the Web: evidence from business-to-consumer portals, **Journal of Services Marketing**, v.17 n.4/5, p.529–547, 2003.

GRÖNROOS, C. A Service Quality Model and Its Marketing Implications, **European Journal of Marketing**, v.18, n.4, p.36-44, 1984.

GUMMERUS, J., LILJANDER, V., PURA, M. van RIEL, A., Customer loyalty to content-based web sites: the case of an online health-care service, **Journal of Services Marketing**, v.18. n.3, p.175-86, 2004.

GÜNTHER, H. **Como elaborar um questionário** (Série: Planejamento de Pesquisa nas Ciências Sociais, nº 01). Brasília, DF: UnB, Laboratório de Psicologia Ambiental, 2003.

HAGA, W.J., ZVIRAM, M. Information systems effectiveness: research designs for causal inference, **Journal of Information Systems**, v.4, n.2 p.141-166, 1994.

- HAIR, J. F., BABIN, B., MONEY, A. H., SAMOUEL, P. **Fundamentos de métodos de pesquisa em administração**. Porto Alegre: Bookman, 2005. 471p.
- HAIR, J. F., BLACK, W. C., BABIN, B.J., ANDERSON, R.E., TATHAM, R.L. **Análise multivariada de dados**. 6. ed. Porto Alegre: Bookman, 2009
- HAMBLETON, R.K., SWAMINATHAN, H., **Item response theory**: Principles and applications. Norwell, MA: Kluwer Academic Publishers, 1985.
- HAMBLETON, R.K., SWAMINATHAN, H., ROGERS, H.J. **Fundamentals of item response theory**. Newbury Park, CA: Sage, 1991.
- HAMILTON, J. Benchmarking on-line services industries, **Journal of Systems Science and Systems Engineering**, v.15, n.1, p.48-69, 2006.
- HART, C. (1999) **Doing a Literature Review**: Releasing the Social Science Research Imagination, (1st) Sage Publications Ltd
- HARTIG, J., HÖHLER, J. Multidimensional IRT models for the assessment of competencies. **Studies in Educational Evaluation**, v.35, p.57-63, 2009.
- HARTIG, J., HÖHLER, J. Representation of competencies in multidimensional IRT models with within-item and between-item multidimensionality. **Journal of Psychology**, v.216, p.88-100, 2008.
- HASAN L., ABUEL RUB E. Assessing the quality of web sites, **Applied Computing and Informatics**, v.9 n.1, p. 11-29, 2012.
- HASSENZAHL, M., The interplay of beauty, goodness, and usability in interactive products. **Human-Computer Interaction**, v.19 n.4 p.319 – 349, 2004.
- van der HEIJDEN, H. Factors influencing the usage of websites: the case of a generic portal in the Netherlands, **Information & Management**, n.40, p. 541–549, 2003.
- HERNÁNDEZ, B., JIMÉNEZ, J., MARTÍN M. J. Key website factors in e-business strategy, **International Journal of Information Management** v.29, p 362–371, 2009.
- HO C.-I., LEE, Y.-L. The development of an e-travel service quality scale, **Tourism Management**, v.26, p. 1434–1449, 2007.
- HOLSTI, O. R. **Content Analysis for the Social Sciences and Humanities**: Reading, Mass, 1969.
- HOLZINGER, K. J., SWINEFORD, F. The bi-factor method. **Psychometrika**, v.2, p.41-54, 1937.
- HORAN, P. **Developing an effectiveness evaluation framework for destination management systems**. Ph.D. thesis. Queen Margaret University, Edinburgh, September, 2010.

HORST, P. **Factor Analysis of Data Matrices**. New York: Holt, Rinehart and Winston, 1965.

HOUTS, C. R., CAI, L.. **flexMIRTMM users manual version 1.0**: Flexible multilevel item factor analysis and test scoring. Seattle, WA: Vector Psychometric Group, 2012.

HHS - U.S. Department of Health and Human Services . **The Research-Based Web Design & Usability Guidelines** (Guidelines). 2006 disponível em: <http://www.usability.gov/guidelines/index.html>. Acesso em 15 de maio 2010.

HUANG, Z. A comprehensive analysis of U.S. counties' e-Government portals: development status and functionalities, **European Journal of Information Systems**, v.16, p.149–164, 2007.

HUIZINGH, E.K. The content and design of web sites: an empirical study. **Information & Management**, v.37, n. 3, p. 123–134, 2000.

HUNG, W.-H., McQUEEN, R.J. Developing an Evaluation Instrument for eCommerce Web Sites from the First-Time Buyer's Viewpoint, **Electronic Journal of Information Systems Evaluation**, v.7, n.1, p31-42, 2004.

GUMMERUS, J., LILJANDER, V., PURA, M., van RIEL, A., Customer loyalty to content-based web sites: the case of an online health-care service, **Journal of Services Marketing**, v.18, n.3, p. 175-86, 2004.

GUO, Y. M., POOLE, M. S., Antecedents of flow in online shopping: a test of alternative models, **Information Systems Journal**, v.19, p.369–390, 2009.

GÜNTHER, H. **Como elaborar um questionário** (Série: Planejamento de Pesquisa nas Ciências Sociais, nº 01). Brasília, DF: UnB, Laboratório de Psicologia Ambiental, 2003.

IBRAHIM, E.E., JOSEPH M., IBEH, K.I.N. Customers' perception of electronic service delivery in the UK retail banking sector, **International Journal of Bank Marketing**, v.24, n.7,p.475–493, 2006.

IEEE Std. 2001–2002, **Recommended Practice for the Internet Web Site Engineering**, Web Site Management, and Web Site Life Cycle, IEEE, 2003.

**INTERNET WORLDS STATS** (2011) <http://www.internetworldstats.com/> Acessado em 17 de setembro de 2012.

INKPEN, K., DEARMAN, D., ARGUE, R. Left-Handed Scrolling for Pen-Based Devices. **International Journal of Human-Computer Interaction**, v.21, n.1, p.91-108, 2006.

ISO9126 **Information Technology** — Software Product Evaluation — Quality Characteristics and Guidelines for Their Use, International Organisation for Standardization, Geneva, 1992.

- JANDA, S., TROCCHIA, P. J., GWINNER, K. P., Consumer perceptions of Internet retail service quality. **International Journal of Service Industry Management**, v.13, n.5, p.412–431, 2002.
- JENNRICH, R.I., BENTLER, P.M. Exploratory bi-factor analysis. **Psychometrika**, v.76, p.537– 549, 2011.
- JOE, H., MAYDEU-OLIVARES, A. A general family of limited information goodness-of-fit statistics for multinomial data. **Psychometrika**, v.75, p.393–419, 2010.
- JOHNSON, R., WICHERN, D. **Applied Multivariate Analysis**, Prentice Hall, 6ª edição, New Jersey, 2007.
- JUN, M. YANG Z., KIM, D. Customers' perceptions of online retailing service quality and their satisfaction, **International Journal of Quality and Reliability Management**, v.21, n.8, p.817–840, 2004.
- JUNKER, B. W. Conditional association, essential independence and monotone unidimensional item response models. **The Annals of Statistics**, v.3, p.1359 –1378, 1993.
- KAISER, H. F. The varimax criterion for analytic rotation in factor analysis. **Psychometrika**, v.23, p.187–200, 1958.
- KALBACH, J. I'm feeling lucky: The role of Emotions in seeking information on the Web. **Journal of the American Society for Information Science and Technology**, v.57, n.6, p.813–818, 2006.
- KAPLAN, D. KRISHNAN, R. PADMAN, R. PETERS, J. Assessing data quality in accounting information systems. **Communications of the ACM**, v.41, n.2, p. 72–78, 1998.
- KARVONEN, K., **The beauty of simplicity**. In: Proceedings of the ACM Conference on Universal Usability (CUU 2000), Washington DC, USA, november, 2000.
- KELDERMAN, H., Loglinear multidimensional item response models. In W. J. van der Linden and Hambleton (Ed.), **Handbook of Modern Item Response Theory** (pp. 287–321). New York: Springer, 1997.
- KELDERMAN, H., RIJKES, C. P. M. Loglinear multidimensional IRT models for polytomously scored items. **Psychometrika**, v.59, p.149–176, 1994.
- KETTINGER, W.J., LEE, C.C. Perceived service quality and user satisfaction with the information services function, **Decision Sciences**, v.25, p. 737–766, 1994.
- KIM Changsu, GALLIERS Robert D., SHIN Namchul, RYOO Joo-Han, KIM Jongheon, Factors influencing Internet shopping value and customer repurchase intention, **Electronic Commerce Research and Applications**, v.11, n.4, p. 374–387, 2012.

KIM, J., LEE J., D. CHOI, Designing emotionally evocative homepages: an empirical study of the quantitative relations between design factors and emotional dimensions, **International Journal of Human-Computer Studies**, v.59, n.6, p. 899–940, 2003.

KIM, D. J., SONG,Y. I., BRAYNOV, S.B., RAO, H.R. A multidimensional trust formation model in B-to-C e-commerce: a conceptual framework and content analyses of academia/ practitioner perspectives. **Decision Support Systems**, v.40, p.143–165, 2005.

KIM, S., STOEL, L. Dimensional hierarchy of retail website quality. **Information & Management**, v.41, n.5, p.619–633, 2004.

KING, W.R. EPSTEIN, B. Assessing information system value: An experimental study. **Decision Sciences**, v.14, n.1, p. 34–45,1983.

KITCHENHAM, B., PFLEEGER, S. L. Software quality: the elusive target. **IEEE Software**, v.13, n.1, p.12-21, 1996.

KLINE, P. **The handbook of psychological testing**, Routledge, London , 2000.

KOYANI, S. J., BAILEY, R. W., NALL, J. R. **Research-based web design and usability guidelines**. US Department of Health and Human Services, 2006. Disponível em <http://usability.gov/pdfs/guidelines.html> Acessado 30 de agosto de 2010.

KAISER, H.F. The varimax criteria for analytical rotation in factor analysis. **Psychometrika**, v.23 p.141-51, 1958.

KULVIWAT, S. BRUNER II, G.C. KUMAR, A. NASCO, S.A. CLARK, T. Toward a unified theory of consumer acceptance of technology, **Psychology & Marketing**, v. 24, n.12, p.1067–1092, 2007.

KUROSU, M., KASHIMURA. K., **Apparent usability vs. inherent usability**: experimental analysis on the determinants of the apparent usability. In CHI, Denver, Colorado, United States, ACM Press, p. 292 – 293, 1995.

LADHARI, R. Developing e-service quality scales: A literature review. **Journal of Retailing and Consumer Services**. v. 17, v.6, p.464-477, 2010.

LAWLEY, D. N. On problems connected with item selection and test construction. **Proceedings of the Royal Society of Edinburgh**, v.61, p. 273-287, 1943.

LEDESMA, R. D., VALERO-MORA, P.. Determining the number of factors to retain in EFA: An easy-to-use computer program for carrying out parallel analysis. **Practical Assessment, Research and Evaluation**, v.12, p.1-11, 2007.

LEE G., LIN, H. Customer perceptions of e-service quality in online shopping, **International Journal of Retail and Distribution Management**, v.33, n.2, p.161–176, 2005.



- LEE, Y., KOZAR, K. A. Understanding of website usability: Specifying and measuring constructs and their relationships. **Decision Support Systems**, v.52 n.2, p. 450–463, 2012.
- LINDROOS, K. Use Quality and the World Wide Web. **Information and Software Technology**, v.39, n.12, p. 827-836, 1997.
- LI, Y., BOLT, D. M., FU, J. A comparison of alternative models for testlets. **Applied Psychological Measurement**, v.30, p.3–21, 2006.
- LI, Y.; LISSITZ, R. W.Exploring the full-information bifactor model in vertical scaling with construct shift, **Applied Psychological Measurement**, v.36 n.1 p.3-20 Jan 2012
- LI, Y., RUPP A. A. Performance of the  $S - \delta^2$  Statistic for Full-Information Bifactor Models. **Educational and Psychological Measurement**, v.71, n.2, p.1-20, 2011.
- LI, Y., TAN, K., XIE, M. Measuring web-based service quality. **Total Quality Management**, v.13, n.5, p.685-70, 2002.
- LINDGAARD, G., Aesthetics, visual appeal, usability, and user satisfaction: What do the user's eyes tell the user's brain? **Australian Journal of Emerging Technologies and Society**, v.5, n.1, p.1-14, 2007.
- LINDGAARD, G., GARY J. F., CATHY D., J. BROWN. **Attention Web Designers: You Have 50 Milliseconds to Make a Good First Impression!** Behavior and Information Technology v.25 p115-26, 2006.
- LINDROOS, K., Use Quality and the World Wide Web, **Information and Software Technology**, v.39, p. 827-836, 1997.
- LIU, C., ARNETT, K.P., Exploring the factors associated with web site success in the context of electronic commerce, **Information and Management**, v.38 n.1, p. 23-33, 2000.
- LIU C.T., DUB T. C., TSAI H. H. A study of the service quality of general portals, **Information & Management**, v.46, n.1, p.52-56, 2009.
- LLOSA, S., J.-L. CHANDON, C. ORSINGER. An Empirical Study of SERVQUAL's Dimensionality, **The Service Industries Journal** v.18, v.2, p. 16-44, 1998.
- LOIACONO, E., WATSON, R., GOODHUE, D., **WEBQUAL: a measure of web site quality**, in Evans, K., Scheer, L. (Eds),Marketing Educators' Conference: Marketing Theory and Applications, v. 13 p.432-7, 2002.
- LONG, M., McMELLON, C. Exploring the determinants of retail service quality on the internet, **Journal of Services Marketing**, v.18, n.1, p. 78–90, 2004.

LORD, F. M. **Applications of item response theory to practical testing problems**. Hillsdale NJ: Erlbaum, 1980.

LORD, F. M. A theory of test scores. Psychometric Monograph No. 7, **Psychometric Society**, 1952.

LORD, F. M. Estimation of latent ability and item parameters when there are omitted responses. **Psychometrika**, v.39, p247-264. 1974.

LORD, F. M.; NOVICK, M. R. **Statistical Theories of Mental Test Score**. Reading: Addison-Wesley, 1968.

MA, H., MENG, C., XIAO, J. The development strategy of electronic commerce in China: New perspective and policy implications, **Journal of Science and Technology Policy in China**, v.1 n.2, p.135 – 147, 2010.

MADU, C.N., MADU, A.A., Dimensions of e-quality, **International Journal of Quality & Reliability Management**, v.19 n.3, p. 246-58, 2002.

MASTERS, G.N., A Rasch model for partial credit scoring. **Psychometrika**, v.47, p.149-174, 1982.

MASTERS G.N., WRIGHT, B.D., **The partial credit model**. In W.J. van der Linden, R.K. Hambleton (Eds) Handbook of modern item response theory. Springer, New York.

MBIPOM, G. HARPER, S. **Visual Aesthetics and Accessibility**: Extent and Overlap. HCW— EIVAA Technical Report 2, February, 2009.

McDONALD, R. P. A basis for multidimensional item response theory. **Applied Psychological Measurement**, v.24, p.99–114, 2000.

McDONALD, R. P. **Nonlinear factor analysis** (Psychometric Monographs, No. 15). Iowa City: Psychometric Society, 1967.

MENDES, E. L. **Uma metodologia para avaliação da satisfação do consumidor com os serviços prestados pelas distribuidoras de energia elétrica**. Rio de Janeiro, 2006. 148p. Tese de Doutorado – Departamento de Engenharia Elétrica, Pontifícia Universidade Católica do Rio de Janeiro.

MILONE, G. **Estatística Geral e Aplicada**. São Paulo: Pioneira Thomson Learning, 2004.

MIN, K.-S. **The Impact of Scale Dilation on the Quality of the Linking of Multidimensional Item Response Theory Calibrations**. PhD Dissertation, Michigan State University, East Lansing, MI, 2003.

MISHA W. V., A. DILLON, Why structure and genre matter for users of digital information: A longitudinal experiment with readers of a web-based newspaper, **International Journal of Human-Computer Studies**, v.64 n.6, p.502-26, 2006.

MISLEVY, D. J., BOCK, R. D. **BILOG**: Item analysis and test scoring with binary logistic models [Computer program]. Chicago: Scientific Software, 1990.

MONK, A., LELOS, K.. **Changing only the aesthetic features of a product can affect its apparent usability**. in International Federation for Information Processing (IFIP), Home informatics and telematics: ICT for the next billion, eds Venkatesh, A., Gonzalez, T., Monk, A., Buckner, B. (Boston:Springer), 241:221–233, 2007.

MOTTA, P. C. **Serviços**: pesquisando a satisfação do consumidor. Rio de Janeiro: Imprinta Express, 1999.

McCLOSKEY, D. Evaluating electronic commerce acceptance with the technology acceptance model, **Journal of Computer Information Systems**, v. 44, n.2, p. 49–57, 2004.

McDONALD, R.P., The Dimensionality of test and items. **The British Journal of Mathematical and Statistical Psychology**, n. 33, p. 161–183, 1981.

McKINNEY, L.N. Creating a Satisfying Internet Shopping Experience via Atmospheric Variables. **International Journal of Consumer Studies**, v.28, n.3, p.268–283, 2004.

MOSIER, C. I. A modification of the method of successive intervals. **Psychometrika**, v. 7 n. 1 p. 19–29, 1942.

MOSIER, C. I. A psychometric study of meaning. **Journal of Social Psychology**, v.13, p.123–140. 1941.

MONSUWE, P. T., DELLAERT, B., G., C., RUYTER, K. What drives consumers to shop online? A literature review. **International Journal of Service Industry Management**, v.12, n.1, p.102–121, 2004.

MURAKI, E. A generalized partial credit model: application of an EM algorithm. **Applied Psychological Measurement**, v.16, p.159–176, 1992.

MURAKI E., CARLSON, J. E. Full-Information Factor Analysis for Polytomous Item Responses, **Applied Psychological Measurement**, v.19, n.1, p.73–90, 1995.

MURAKI, E., ENGELHARD, G. Full Information Item Factor Analysis: applications of EAP scores. **Applied Psychological Measurement**, v.9, p.417–430, 1985.

MUTHÉN, B. Contributions to factor analysis of dichotomous variables. **Psychometrika**, v.43, p.551–560, 1978.

NANDAKUMAR, R. Traditional Dimensionality Versus Essential Dimensionality, **Journal of Educational Measurement**, v.28, n. 2, p. 99–117, 1991.

NUSAIR, K., KANDAMPULLY, J., The antecedents of customer satisfaction with online travel services: a conceptual model, **European Business Review**, v.20 n.1, p. 4-19, 2008.

NEILL, J. **Qualitative versus Quantitative Research: Key Points in a Classic Debate**, 2007. Disponível em <http://wilderdom.com/research/QualitativeVersusQuantitativeResearch.html>, acessado em 04/04/2010

NELSON, K.G. Global information systems quality: key issues and challenges. **Journal of Global Information Management**, v.4, n.4, p. 4–14, 1996,.

NIELSEN, J., **Web 2.0 can be dangerous**. Available from: <<http://www.useit.com/alertbox/web-2.html>>, 2007 (acessado em 13/7/2010).

NYLUND, K. L., ASPAROUHOV, T., MUTHÉN, B. Deciding on the number of classes in latent class analysis and growth mixture modeling. A Monte Carlo simulation study. **Structural Equation Modeling**, v.14, p.535–569, 2007.

NUSAIR, K., KANDAMPULLY, J., The antecedents of customer satisfaction with online travel services: a conceptual model, **European Business Review**, v.20, n.1, p.4-19, 2008.

O'CONNOR, T. **Lecture notes for research methods**, 2007. Disponível em <http://aulettesociology.tripod.com/id20.html> Acessado em 10 abril 2011.

ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT. PISA 2003 Technical Report. Paris: OECD, 2005.

O'NEILL, M. WRIGHT, C. FITZ, F. Quality evaluation in on-line service environments: an application of the importance–performance measurement technique, **Managing Service Quality**, v.11 n.6, p. 402–417, 2001.

OSHIMA, T. C., MILLER, M. D. Multidimensionality and item bias in item response theory. **Applied Psychological Measurement**, v.16, p.237-248, 1992.

OLSINA, L., GODOY, D., LAFUENTE, G.J., ROSSI, G. **Specifying Quality Characteristics and Attributes for Websites**, in Proceedings of the First ICSE Workshop on Web Engineering, 16 -17 May Los Angeles, USA, 1999.

OVERBEEKE, K., WENSVEEN, S. Beauty in use. **Human-Computer Interaction**, v.19 n.4, p.367–369, 2004.

PADOVANI, S., MOURA, D. **Navegação em hipermídia: uma abordagem centrada no usuário**. Editora Ciência Moderna Ltda, Rio de Janeiro – RJ, 2008.

PALMER J.W., Web site usability, design, and performance metrics, **Information Systems Research** v.13, n.2, p. 151–67, 2002.

- PANDIR, M. KNIGHT. J. Homepage aesthetics: The search for preference factors and the challenges of subjectivity, **Interacting with Computers** v.18, p.1351–1370, 2006.
- PARASURAMAN, A., ZEITHAML, V. BERRY, L. SERVQUAL: A multi-item scale for measuring consumer perception of service quality. **Journal of Retailing**, v.64, n.1, p.12-40, 1988.
- PARASURAMAN, A., ZEITHAML, V., Leonard L. B., A Conceptual Model of Service Quality and Its Implications for Future Research, **Journal of Marketing**, v.49, p.41-50, 1985.
- PARASURAMAN, A., ZEITHAML, V.A., MALHOTRA, A. E-S-Qual: a multiple-item scale for assessing electronic service quality, **Journal of Service Research**, v.7 n.3, p. 213–233, 2005.
- PASQUALI, L. Princípios de elaboração de escalas psicológicas, **Revista de Psiquiatria Clínica**, v.25, n.5, 1998.
- PASQUALI, L. **Psicometria**: teoria e aplicações. Brasília : Editora Universidade de Brasília, 1997.
- PETRE, M., S. MINOCHA, D. ROBERTS. Usability Beyond the Website: an Empirically Grounded E-commerce Evaluation Instrument for the Total Customer Experience. **Behaviour & Information Technology**, v.25, n.2, p.189–203, 2006.
- PINTERITS, A. TREIBLMAIER, H. POLLACH I. Environmental websites: an empirical investigation of functionality and accessibility. **International Journal of Technology, Policy and Management**, v.6, n.1, 2006.
- PODSAKOFF, P. M., AHEARNE, M., MACKENZIE, S. B. Organizational citizenship behaviors and the quantity and quality of work group performance. **Journal of Applied Psychology**, v.82, p.262-270, 1997.
- R Core Team (2012). **R**: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org/>.
- RANGANATHAN, C., GANAPATHY, S. Key dimensions of business-to-consumer web sites. **Information & Management**, v.39, n.6, p. 457-465, 2002.
- RASCH, G. **Probabilistic Models for Some Intelligence and Attainment Tests**. Copenhagen, Denmark: Danish Institute for Educational Research. 1960.
- RAUCH, D., HARTIG, J. Multiple-Choice versus Open-Ended Response Formats of Reading Test Items: A Two-Dimensional IRT Analysis. **Psychological Test and Assessment Modeling**, v.52, n.4, p.354-379, 2010.
- RECKASE, M. D. **Multidimensional Item Response Theory**, Springer, New York – USA, 2009.

RECKASE, M. A linear logistic multidimensional model for dichotomous items response data. In W. J. van der Linden and Hambleton (Ed.), *Handbook of Modern Item Response Theory* (pp. 271-286). New York: Springer, 1997.

RECKASE, M. D., McKINLEY, R. L. The discriminating power of items that measure more than one dimension. **Applied Psychological Measurement**, v.15, p.361-373, 1991.

REISE S.P., MORIZOT J, HAYS R.D. The role of bifactor model in resolving dimensionality issues in health outcome measures. **Quality of Life Research**, v.16, p.19-31, 2007.

RICHARDSON M.W. The relationship between difficulty and the differential validity of a test. **Psychometrika**. v.1 p.33-49. 1936.

RIJMEN, F., Formal relations and an empirical comparison among the bi-factor, the testlet, and a second-order multidimensional IRT model. **Journal of Educational Measurement**, v.47, p.361-372, 2010.

RIJMEN, F., De BOECK, P. A relation between a between-item multidimensional IRT model and the mixture-Rasch model. **Psychometrika**, v.70, p.481-496, 2005.

ROBERTS, J. S., DONOGHUE, J. R., LAUGHLIN, J. E. A general item response theory model for unfolding unidimensional polytomous responses. **Applied Psychological Measurement**, v.24, n.1, p.3-32, 2000.

ROBERTS, J.S., HUANG, C.W. GGUMLINK: A computer program to link parameter estimates of the generalized graded unfolding model from item response theory. **Behavior Research Methods**, v.35, n.4, p.525-536, 2003.

ROHATGI, V. K., SALEH, E. A.K.M. **An introduction to probability and statistics**. Wiley Series in Probability and Statistics, Chichester, 2001.

ROMEU, J. L. On Operations Research and Statistic Techniques: Keys to Quantitative Data Mining, **American Journal of Mathematical and Management Sciences**, v. 26 n.3, p 293-32, 2007.

ROSSITIER, J.R. ER-SERVCOMPSQUAL: A Measure of E-Retailing Service Components Quality. **Service Science**, v.1, n.4, p.212-22, 2009.

ROST, J. Logistic Mixture Models. In W. J. van der Linden and Hambleton (Ed.), *Handbook of Modern Item Response Theory* (pp. 449-463). New York: Springer, 1997.

ROTH, S. SCHMUTZ, P. PAUWELS, S. BARGAS-AVILA, J. OPWIS K.. Mental models for web objects: where do users expect to find the most frequent objects in online shops, news portals, and company web pages? **Interacting with Computers**, v.22 n.2, p. 140-15, 2010.

ROY, B. **Méthologie multicritère d'aide à la decision**. Economica, Paris, 1985.

- RUST, R. T., The Rise of E-Service, **Journal of Service Research**, v.3, p.283-84, 2001.
- SAMEJINA, F. Estimation of latent ability using a response pattern of graded scores. **Psychometrika** Monograph Supplement, 1, 1969.
- SANTOS, A. R.. **Metodologia científica**: a construção do conhecimento. Rio de Janeiro: DP & A, 1999.
- SANTOS, J., e-Service quality: a model of virtual service quality dimensions, **Managing Service Quality**, v.13, n.3, p.233-46, 2003.
- SCHARL, A., POLLACH, I., BAUER, C. Determining the Semantic Orientation of Web-Based Corpora, **Intelligent Data Engineering and Automated Learning**, v.2690, p.840-849, 2003,
- SCHENKMAN B.N., JONSSON, F.U. Aesthetics and preferences of web pages. **Behavior and Information Technology**, v.19 n.5, p. 367–377, 2000.
- SCHÄRLIG, A. **Décider sur plusieurs critères**. Collection Diriger l'Enterprise, Press Polytechniques Romandes, 1985.
- SCHMETTOW, M., VIETZE, W., **Introducing item response theory for measuring usability inspection processes**, Proceeding of the twenty-sixth annual SIGCHI conference on Human factors in computing systems, April 05-10, Florence, Italy, 2008.
- SCHNEIDEWIND, N. F. Methodology for validating software metrics, **IEEE Transaction Software Engineering**, v.18, n.5, p.410-422, 1992.
- SCHUBERT, P., DETTLING, W. **Extended Web assessment method (EWAM)-evaluation of e-commerce applications from the customer's viewpoint**. Hawaii International Conference on Systems Science, Hawaii 2002.
- SCHUMACKER, R. E. **Item Response Theory**. Applied Measurement Associates, 2005, disponível em <http://www.appliedmeasurementassociates.com/articles/ITEM%20RESPONSE%20THEORY.pdf> acessado em 10 março 2011.
- SCHUMAN, H. KALTON, G. **Survey methods**. In G. Lindezey; E. Aronson (Eds.), Handbook of social psychology, 3ªed.. V. 1 New York: Random House, p 635-697, 1985
- SHACHAF, P., OLTMANN, S.M. HOROWITZ, S.M., Service equality in virtual reference, **Journal of the American Society for Information Science and Technology**, v.59, n.4, p.535-50, 2008.
- SEFFAH, A., DONYAEE, M., KLINE, R. B., HARKIRAT K., PADDA,K. Usability measurement and metrics: A consolidated model, **Software Quality Control**, v.14 n.2, p.159-178, 2006.

SEO, D. G., Application of the Bifactor Model to Computerized Adaptive Testing, Ph.D. thesis. University of Minnesota, January 2011.

SERKAN, A., EDA, A., SAFAK, A. Re-assessment of E-S-Qual and E-RecS-Qual in a pure service setting. **Journal Business Research**, v.63, n.3, p.232-40, 2010.

SHACHAF, P., OLTMANN, S.M., HOROWITZ, S.M., Service equality in virtual reference, **Journal of the American Society for Information Science and Technology**, v.59, n.4, p. 535-50, 2008.

SPECTOR, P. E., VAN KATWYK, P. T., BRANNICK, M. T., CHEN, P. Y. When two factors don't reflect two constructs:How item characteristics can produce artificial factors. **Journal of Management**, v.23, p.659-677, 1997.

CHO, S.-J., ATHAY, M., PREACHER, K. J. (in press). Measuring change for a multidimensional test using a generalized explanatory longitudinal item response model, **British Journal of Mathematical and Statistical Psychology**

SIEBER, S., SABATIER, J.V. Market bundling strategies in the horizontal portal industry, **International Journal of Electronic Commerce**, v.7, n.4, p.37-54, 2003.

SILVA, E. L., MENEZES, E. M. **Metodologia da pesquisa e elaboração de dissertação**. Florianópolis: UFSC/PPGEP/LED, 2000, 118p.

SIRECI, S. G., THISSEN, D., WAINER, H. On the reliability of testlet-based tests. **Journal of Educational Measurement**, v.28, p.237-247, 1991.

SMITH, A. G. Applying evaluation criteria to New Zealand government Websites. **International Journal of Information Management** v.21, p 137-149, 2001.

SOARES, T.M. Utilização da Teoria da Resposta ao Item na Produção de Indicadores Sócio-Econômicos. **Pesquisa Operacional**, v.25, n. 1, p.83-112, 2005.

SOHN C., TADISINA, S.K., Development of e-service quality measure for internet-based financial institutions, **Total Quality Management and Business Excellence** v.19, n.9, p.903-918, 2008.

SONG, J., F. M. ZAHEDI. **Web-design in e-commerce: A theory and empirical analysis**. Veda Storey, Sumit Sarkar, Janice I. De Gross, eds. Proc. 22nd Internat. Conf. Inform. Systems, New Orleans, LA, p205-219, 2001.

SOMMER, B; SOMMER, R. **A practical guide to behavioral research: Tools and techniques** (4th ed). New York: Oxford U Press.

SOWA, J.F. **Conceptual Structures**, Reading, Addison-Wesley Pub., MA, 1984.

SOWA, J.F. **Principles of Semantic Networks: Explorations in the Representation of Knowledge**, Morgan Kaufmann Publishers, San Mateo, CA, 1991.

SPENCER, S. G. **The strength of multidimensional item response theory in exploring construct space that is multidimensional and correlated**. Unpublished doctoral dissertation. Brigham Young University, USA, 2004.



STEPCHENKOVA S., et al. Benchmarking CVB website performance: Spatial and structural patterns, **Tourism Management**, v.31, p.611–620, 2010.

STEVERSON, W. **Estatística aplicada à administração**. São Paulo: Harper & Row do Brasil, 1981.

STIAKAKIS, E., C.K. GEORGIADIS. E-Service Quality: Comparing the Perceptions of Providers and Customers. **Managing Service Quality**, v.19, n.4 p.410-430, 2009.

STOREY, M.-A., PHILLIPS, B., MACZEWSKI, M., WANG, M. Evaluating the usability of Web-based learning tools. **Educational Technology & Society**, v.5 n.3 2002

STOUT, W. F. A new item response theory modelling approach with applications to unidimensionality assessment and ability estimation. **Psychometrika**, v.55, p.293-326, 1990.

TATE, R. A comparison of selected empirical methods for assessing the structure of binary data test. **Applied Psychological Measurement**, v.27, p.159–203, 2003.

TAYLOR S. J., BOGDAN, R. **Introduction to Qualitative Research Methods: a guidebook and resource** New York, NY, USA, p.24-43, 1997.

TAYLOR, M.J., ENGLAND, D. (2006), Internet marketing: web site navigational design issues, **Marketing Intelligence & Planning**, v.24, n.1, p.77-85, 2006.

TEDESHCI, B. E-commerce report; log in your measurements, and the clothes may well fit, **The New York Times**, 12, p. 10, March 2007.

TEZZA, R. **Proposta de um construto para medir usabilidade em site de e-commerce utilizando a Teoria da Resposta ao Item**. Dissertação de Mestrado - Programa de Pós-Graduação em Engenharia de Produção, UFSC, Florianópolis, 2009.

TEZZA, R., BORNIA, A.C., ANDRADE, D.F., Measuring web usability using item response theory: Principles, features and opportunities, **Interacting with Computers**, v. 23, n.2, p.167–175, 2011.

THISSEN, D., STEINBERG, L. A response model for multiple choice items. **Psychometrika**, v.49, n.4 p. 501-519, 1984.

THURSTONE, L. L. A law of comparative judgments. **Psychological Review**, v.34, p.278–286, 1928.

THURSTONE, L. L. Attitudes can be measured. **American Journal of Sociology**, v.33, p.529–554, 1927.

THURSTONE, L.L. **Multiple factor analysis**. Chicago, University of Chicago Press, 1947.

TOWNSEND, M., KONOLD, T. R.. Measuring early literacy skills: A latent variable investigation of the Phonological Awareness Literacy Screening for Preschool. **Journal of Psychoeducational Assessment**, v.28, p.115-128, 2010.

TRACTINSKY, N. SHOVAL-KATZ A., IKAR, D. What is beautiful is usable, **Interacting with Computers** v.13, p. 127–145, 2000.

TRAUB, R. E. **A priori considerations in choosing an item response model**. In R. K. Hambleton (Ed.), Applications of item response theory. Vancouver BC: Educational Research Institute of British Columbia, 1983.

TREIBLMAIER, H. Web site analysis: A review and assessment of previous research, **Communications of the AIS**, v.19, p.806-843, 2007.

TURLEY, L. W., MILLIMAN, R. E. Atmospheric effect on shopping behavior: A review of the experimental evidence. **Journal of Business Research**, v.49, p.193–211, 2000.

van der LINDEN, W.J., HAMBLETON, R.K., **Handbook of Modern Item Response Theory**. New York: Spring- Verlag, 1997.

van der MERWE, R., BEKKER, J. A framework and methodology for evaluating e-commerce Web sites. **Internet Research: Electronic Networking Applications and Policy**, v.13 n.5 p.330-341, 2003.

VENKATESH, V., R. AGARWAL. Turning Visitors into Customers: A Usability-Centric Perspective on Purchase Behavior in Electronic Channels. **Management Science** v.52 n.3, p.367-382, 2006.

VILLAS, M. V.; VAN ADUARD MACEDO-SOARES, T. D. L.; RUSSO, G. M. Bibliographical research method for business administration studies: a model based on scientific journal ranking. **Brazilian Administration Review**, v. 5, n. 2, p. 139-159, 2008.

VINCKE, PHILIPPE. **Multicriteria Decision-Aid**. New York: Wiley, 1992

WAINER, H. BRADLOW, E.T., WANG, X. **Testlet Response Theory and its Applications**, Cambridge University Press, New York, USA, 2007

WALKER, C.M. BEREBTAS, S.N., Comparing Multidimensional and Unidimensional Proficiency Classifications: Multidimensional IRT as a Diagnostic Aid. **Journal of Educational Measurement**; Fall 2003, v. 40 n.3, p.255-276, 2003.

WANG, P., GAO F., Full-Information Item Bifactor Analysis of the Job Burnout Scale for Chinese College Teachers, **Journal of Convergence Information Technology**, v. 5, n.7, p.155-160, 2010.

WANG, R., STOREY, V., FIRTH, A. A framework for data quality research. **IEEE Transactions on Knowledge and Data Engineering**, v.7, n.4 , p. 623–640, 1995.

WILSON, M. **Constructing Measures**: na item response modeling approach, Lawrence Erlbaum Associates, Publishers, Mahwah, New Jersey, London, 2005.

WILSON, M., HOSKENS, M. **Multidimensional item responses**: Multimethod-multitrait perspectives. In S. Alagumalai et al. (Eds.), *Applied Rasch measurement: A book of exemplars*.(pp. 287-307). New York: Springer, 2005.

WOELFEL, J. **Neural Networks**: Applications of the Cognitive Revolution to Advertising, Marketing, and Political Research, Terra Research and Computing, Troy, NY, 1990.

WONGTADA, N., RICE, G. Multidimensional latent traits of perceived organizational innovation: Differences between Thai and Egyptian employees. **Asia Pacific Journal of Management**, v.25, p.537–562, 2008.

WOOD, R., WILSON, D., GIBBONS, R., SCHILLING, S., MURAKI, E., BOCK, R. D. (2003). **TESTFACT**: Test scoring, item statistics, and item factor analysis [version 4; computer program]. Lincolnwood, IL: Scientific Software International, Inc. 46.

WU, M. L., ADAMS, R.J., WILSON, M. (1998). **ACERConQuest** [computer program]. Hawthorn, Australia: ACER.

XIE, M. WANG, H. GOH, T.N. Quality dimensions of Internet search engines, **Journal of Information Science**, v.24, n.5, p.365-372, 1998.

XU, J. LIU, Z. **Study of Online Stickiness**: its Antecedents and Effect on Repurchase Intention International Conference on e-Education, e-Business, e-Management and e-Learning. Sanya, China, 2010.

W3C (2008). Web content accessibility guidelines (wcag) 2.0. W3C Recommendation. Disponível em: <http://www.w3.org/TR/WCAG20/>. Acessado em 25 de maio de 2012.

WALKER, C. M., BERETVAS, S. N.. Comparing multidimensional and unidimensional proficiency classifications: Multidimensional IRT as a diagnostic aid. **Journal of Educational Measurement**, v.40, p.255–275, 2003.

WEI, H.. **Multidimensionality in the NAEP science assessment: substantive perspectives, psychometric models, and task design**, PhD Dissertation University of Maryland, 2008.

WILSON, D.T.; WOOD R.; GIBBONS, R.. **Testfact**: Test Scoring, and Item Factor Analysis. Lincolnwood, Illinois: Scientific Software International, Inc., 1998.

WOLFE E.W., HICKEY D.T., KINDFIELD A.C., An application of the multidimensional random coefficients multinomial logit model to evaluating cognitive models of reasoning in genetics. **Journal of Applied Measurement**, v.10 n.2 p.196-207, 2009.

WOLFINBARGER, M. GILLY, M.C. eTailQ: dimensionalizing, measuring and predictingetail quality, **Journal of Retailing**, v.79, n.3, p.183–198, 2003,.

WOOD, R., WILSON, D., GIBBONS, R., SCHILLING, S., MURAKI, E., BOCK, R. D. **TESTFACT**: Test scoring, item statistics, and item factor analysis [version 4; computer program]. Lincolnwood, IL: Scientific Software International, Inc, 2003.

WORWA K., STANIK J. Quality of Web-based information systems, **Journal of Internet Banking and Commerce**, v.15, n.3, p.1-13, 2010.

WATSON, K., et al, Innovative application of a multidimensional item response model in assessing the influence of social desirability, on the pseudo-relationship between self-efficacy and behavior. **Health Education Research, Theory & Practice**, v.21(Supplement 1): i85–i97, 2006.

WROBLEWSKI, L.. **Web Form Design**: filling in the blanks. Rosenfeld Media, NY, 2008.

YANG, Z. CAI, S. ZHOU, Z. ZHOU, N. Development and validation of an instrument to measure user perceived service quality of information presenting Web portals, **Information & Management**, v.42, n.4, p.575–589, 2005.

YANG, Z., JUN,M., Consumer perception of e-service quality: form purchaser and non purchaser perspectives. **Journal of Business Strategies**, v.19, n.1, p.19–41, 2002.

YANG, Z., JUN M., PETERSON, R.T. Measuring customer perceived online service quality: scale development and managerial implications, **International Journal of Operations and Production Management**, v.21, n.11,p. 1149–1174, 2004.

YAO, L., SCHWARZ R. A multidimensional partial credit model with associated item and test statistics: An application to mixed format tests. **Applied Psychological Measurement**, v.30, p.469-492, 2006.

YOO, B., DONTU, N. Developing a scale to measure the perceived quality of Internet shopping sites (SITEQUAL), **Quarterly Journal of Electronic Commerce**, v.2, n.1, p. 31–47, 2001.

ZHANG, P., von DRAN, G.M. Satisfiers and dissatisfiers: a two-factor model for website design and evaluation, **Journal of the American Society for Information Science**, v.51 n.14, p.1253–1268, 2000.

ZHANG, P. von DRAN, G. M. BLAKE, P. PIPITHSUKSUNT, V. Important design features in different website domains: an empirical study of user perceptions, **e-Service Journal**, v.1, n.1, 2001

ZHANG J., STOUT W., Conditional covariance structure of generalized compensatory multidimensional items, **Psychometrika**, v.64, n.2, p.129-152, 1999

ZEITHAML, V.A. PARASURAMAN, A. MALHOTRA, A. **E-service Quality: Definition, Dimensions and Conceptual Model**, Working Paper. Marketing Science Institute, Cambridge, MA

ZEITHAML, V.A. PARASURAMAN, A. MALHOTRA, A. Service quality delivery through websites: a critical review of extant knowledge, **Journal of the Academy of Marketing Science**, v.30, n.4, p. 362–375, 2002.

ZERFASS, A., HARTMANN, B. The Usability Factor: Improving the Quality of E-Content. **EContent**, p.165-182, 2005

## APÊNDICE

### Apêndice 1 – URL dos websites analisados (ordem alfabética).

<http://acessorioscar.com.br/>  
<http://agvimports.webstorelw.com.br>  
<http://amomuito.com/>  
<http://asfilhasdamae.webstorelw.com.br>  
<http://cafepequeno.com.br/>  
<http://ce420friends.webstorelw.com.br/>  
<http://centoeum1.webstorelw.com.br/>  
<http://copymaisonline.com.br/>  
<http://emilemoveis.com.br/>  
<http://frloja.tk/perfumes/>  
<http://hendrixmusic.com.br>  
<http://lenandalingerie.com.br/>  
<http://loja.charutosecachimbos.com.br>  
<http://loja.daikokuorganicos.com.br/>  
<http://loja.itaplantasonline.com.br/>  
<http://loja.tray.com.br/loja/loja-103669>  
<http://lojadece.com.br>  
<http://lojavirtual.lojasrenner.com.br>  
<http://maranatacd.com/>  
<http://mascateonline.com>  
<http://meuvicio.com/>  
<http://minimaxfarma.com.br/>  
<http://mirelafiorese.com/>  
<http://plobera.webstorelw.com.br/>  
<http://tabvirtual.com/>  
<http://uniflores.com.br>  
<http://www.100limitesexshop.com.br/>  
<http://www.2lentes2.com.br/>  
<http://www.7livraria.com.br/>  
<http://www.aazperfumes.com.br/>  
<http://www.absurdofantacias.com.br>  
<http://www.abusecosméticos.com.br>  
<http://www.adegaexpand.com.br>  
<http://www.adesivaria.com/>  
<http://www.adesivosdenotebook.com>  
<http://www.aesportiva.com.br>  
<http://www.africanart.com.br/>  
<http://www.airu.com.br/>  
<http://www.amor-sexshop.com.br>  
<http://www.amyrsom.com.br/>  
<http://www.angeloni.com.br>  
<http://www.animafest.com.br/>  
<http://www.animashopping.com.br/>  
<http://www.apoderosa.com.br/>  
<http://www.aquaricamp.com.br>  
<http://www.aquariushobby.com.br>  
<http://www.araujo.com.br>

<http://www.artemodernamolduras.com.br>  
<http://www.artesanatobresler.com.br/>  
<http://www.asiatronic.com>  
<http://www.atacadaopneus.com.br>  
<http://www.atacadodoprazer.com/>  
<http://www.atera.com.br/>  
<http://www.audiocarcenter.com.br>  
<http://www.autofoxsound.com.br/>  
<http://www.avaistore.com.br/>  
<http://www.bazarpersa.com.br>  
<http://www.beautiful.com.br/loja/>  
<http://www.belaglamour.com.br>  
<http://www.belevita.com.br/>  
<http://www.belinhareplicasdegrife.com.br/>  
<http://www.bellafacil.com.br>  
<http://www.bestgospel.com.br>  
<http://www.birras.com.br>  
<http://www.blessmusic.com.br>  
<http://www.bluetechcomputadores.com.br>  
<http://www.brasiltechinformatica.com.br/>  
<http://www.brinquedoslaura.com.br/>  
<http://www.caculadepneus.com.br>  
<http://www.cafefacil.com.br>  
<http://www.cafestore.com.br>  
<http://www.calvi.com.br/>  
<http://www.camillasflores.com.br>  
<http://www.camisariawf.com/>  
<http://www.camisetasdahora.com/>  
<http://www.caos.com.br>  
<http://www.carambolaonline.com.br>  
<http://www.carsigma.com.br>  
<http://www.casabelem.com.br>  
<http://www.casacenter.com.br>  
<http://www.casadabibliaonline.com>  
<http://www.casasbahia.com.br/>  
<http://www.catalogocdc.com/>  
<http://www.catholicanet.com/>  
<http://www.cazulo.com>  
<http://www.cderistao.com/>  
<http://www.cdpoint.com.br>  
<http://www.centauro.com.br/>  
<http://www.cervejagourmet.com>  
<http://www.cervejastore.com.br>  
<http://www.charutosonline.com.br>  
<http://www.cheirodemusica.com.br>  
<http://www.chicafashion.com.br/>  
<http://www.ciadaslentes.com.br>  
<http://www.ciadoslivros.com.br/>  
<http://www.cinerama.com.br/>  
<http://www.cirurgicazonasul.com.br>  
<http://www.citylar.com.br>  
<http://www.claireandbruce.com.br>  
<http://www.clickjardim.com.br/>

<http://www.clubedomalte.com.br>  
<http://www.coisasdavid.com.br/>  
<http://www.colombo.com.br/>  
<http://www.comperdelivery.com.br>  
<http://www.comprafacil.com.br>  
<http://www.compratesanato.com.br/>  
<http://www.compretv.com/>  
<http://www.cornucopia.com.br/>  
<http://www.corpodahora.com.br/>  
<http://www.corpoperfeito.com.br/>  
<http://www.corpovita.com.br/>  
<http://www.costibebidas.com.br>  
<http://www.cruzterrasanta.com.br/>  
<http://www.dafiti.com.br>  
<http://www.dasvelas.com.br/>  
<http://www.dedobrinquedo.com.br>  
<http://www.dell.com.br>  
<http://www.dellavia.com.br/>  
<http://www.desejosaude.com.br/>  
<http://www.detalhes.net.br>  
<http://www.dfrance.com.br/>  
<http://www.digitalcam.com.br/>  
<http://www.drogariaminasbrasil.com.br>  
<http://www.drogarianovaesperanca.com.br/>  
<http://www.drshapesuplementos.com.br>  
<http://www.drumshop.com.br/>  
<http://www.easy.com.br/>  
<http://www.ecamisetas.com.br/>  
<http://www.eclock.com.br>  
<http://www.ecobiosaude.com.br/>  
<http://www.editorasenacsp.com.br>  
<http://www.elementodecor.com.br/>  
<http://www.e-lens.com.br>  
<http://www.elentes.com.br>  
<http://www.eletroshopping.com.br>  
<http://www.eletrosom.com/>  
<http://www.elevenshop.com.br>  
<http://www.elo7.com.br>  
<http://www.e-lustre.com.br/>  
<http://www.enzom.com.br/>  
<http://www.eotica.com.br>  
<http://www.epocacosmeticos.com.br>  
<http://www.erosmania.com.br/>  
<http://www.erotikatoys.com.br/>  
<http://www.eshopping-me.com>  
<http://www.etna.com.br>  
<http://www.exclusivacastor.com.br>  
<http://www.extra.com.br>  
<http://www.fabricaegaleriabrasil.com.br>  
<http://www.farmaciamix.com.br/>  
<http://www.farmadelivery.com.br>  
<http://www.farmagora.com.br>  
<http://www.fascination.com.br>



<http://www.fastshop.com.br>  
<http://www.festimania.com.br>  
<http://www.flordelizsexshop.com.br/>  
<http://www.floresnaweb.com/>  
<http://www.floresnet.com.br/>  
<http://www.floresonline.com.br>  
<http://www.floriculturadailha.net/>  
<http://www.floripasexstore.com.br/>  
<http://www.fluboutique.com.br>  
<http://www.flyboutiquesensual.com.br/>  
<http://www.fnac.com.br>  
<http://www.foxflores.com/>  
<http://www.freeitalia.com.br/>  
<http://www.fruitfatale.com.br>  
<http://www.fumacaria.com.br>  
<http://www.galeriasp.com.br/>  
<http://www.gangmusic.com.br>  
<http://www.gazin.com.br/>  
<http://www.gbgnpneus.com.br/>  
<http://www.gigantedacolina.com.br>  
<http://www.girafa.com.br/>  
<http://www.giulianaflores.com.br>  
<http://www.gjtecidos.com.br>  
<http://www.gladysreligiosos.com.br/>  
<http://www.glamour.com.br>  
<http://www.gospelgoods.com.br/>  
<http://www.grandcru.com.br>  
<http://www.grannatura.com.br/>  
<http://www.gravuras.com>  
<http://www.gremiomania.com.br/>  
<http://www.groselha.art.br/>  
<http://www.havan.com.br>  
<http://www.havanna.com.br/>  
<http://www.heringwebstore.com.br>  
<http://www.hippo.com.br/>  
<http://www.homepoint.com.br/>  
<http://www.hopelingerie.com.br>  
<http://www.ibitinga.ind.br/aquarela/>  
<http://www.ifragrancia.com/>  
<http://www.imigrantesbebidas.com.br>  
<http://www.induspar.com>  
<http://www.innovainformatica.com.br>  
<http://www.insinuante.com.br>  
<http://www.isexshop.com.br/>  
<http://www.italianorelogios.com>  
<http://www.itboutique.com.br/>  
<http://www.jogmusic.com.br>  
<http://www.joiaslulean.com.br>  
<http://www.jovial.com.br/>  
<http://www.jupiterpet.com.br/>  
<http://www.kabum.com.br>  
<http://www.kanui.com.br/>  
<http://www.kapeh.com.br/lojavirtual/>

<http://www.karicia.com.br/>  
<http://www.kikitostoy.com.br/>  
<http://www.kylixvinhos.com.br/>  
<http://www.lcm.com.br>  
<http://www.leader.com.br/>  
<http://www.leandrinstore.com.br>  
<http://www.leitura.com/>  
<http://www.lenat.com.br/loja/home/>  
<http://www.lentesdecontato.net>  
<http://www.levata.com.br/>  
<http://www.lightinthebox.com>  
<http://www.lily.com.br>  
<http://www.limeart.com.br>  
<http://www.lingerie.com.br/>  
<http://www.lingerishop.com.br/>  
<http://www.livrariabks.com.br/>  
<http://www.livrariacultura.com.br>  
<http://www.livrariadafisica.com.br>  
<http://www.livrariagalileu.com.br>  
<http://www.livrarialuminarium.com.br/>  
<http://www.livrariamartinseditora.com.br/>  
<http://www.livriassampliar.com.br/>  
<http://www.livriassapiens.com.br/>  
<http://www.livriascuritiba.com.br>  
<http://www.livrosdeengenharia.com.br>  
<http://www.lobhi.com.br>  
<http://www.lojabotafogo.com.br>  
<http://www.lojadoflamengo.com.br>  
<http://www.lojadolouvor.com/>  
<http://www.lojadoprazer.com.br/>  
<http://www.lojadosaopaulo.com.br/>  
<http://www.lojadostecidos.com.br/>  
<http://www.lojadosuplemento.com.br/>  
<http://www.lojafluidovital.com.br/>  
<http://www.lojagraodaserra.com.br/>  
<http://www.lojamilenarte.com/>  
<http://www.lojapreferida.com/>  
<http://www.lojascom.com.br>  
<http://www.lojasfurieri.com.br>  
<http://www.lojaskasalar.com.br>  
<http://www.lojaskd.com.br/>  
<http://www.lojasm.com>  
<http://www.lojazuk.com.br/>  
<http://www.londonperfumes.com.br/>  
<http://www.lustresschelmi.com.br/>  
<http://www.luxonopulso.com.br>  
<http://www.luzescenter.com.br/>  
<http://www.macherie.com.br/>  
<http://www.madeinbrazil.com.br>  
<http://www.magazineluiza.com.br>  
<http://www.maisqpneu.com.br>  
<http://www.malaguetasexshop.com.br/>  
<http://www.maniabrinquedos.com.br>

<http://www.maniavirtual.com.br/>  
<http://www.marcynonline.com.br>  
<http://www.maremmoto.com.br/>  
<http://www.mariagabriela.com.br>  
<http://www.maxfesta.com.br>  
<http://www.megalivraria.com>  
<http://www.megamamute.com.br/>  
<http://www.megamonte.com.br>  
<http://www.megasom.com.br>  
<http://www.mensajeiromusical.com.br/>  
<http://www.mercolux.com.br>  
<http://www.meuadesivo.com>  
<http://www.meumoveldemadeira.com.br>  
<http://www.miamirelogios.com.br/>  
<http://www.milaonline.com.br/>  
<http://www.minhacamisetaonline.com/>  
<http://www.missverniz.com.br/>  
<http://www.misterfestas.com.br/>  
<http://www.mmartan.com.br>  
<http://www.momentodovinho.com.br>  
<http://www.morangao.com.br>  
<http://www.movelstore.com.br/>  
<http://www.mpbrinquedos.com.br>  
<http://www.muccashop.com.br/>  
<http://www.multiloja.com.br>  
<http://www.multisom.com.br/>  
<http://www.multivisi.com.br>  
<http://www.multkoisas.com.br>  
<http://www.mundodofutebol.com.br/>  
<http://www.mundomax.com.br>  
<http://www.mundopalmeiras.com.br>  
<http://www.musicalbrasil.com.br/>  
<http://www.musicalgrellmann.com.br>  
<http://www.mytime.com.br>  
<http://www.natureflores.com.br/>  
<http://www.nbfilm.com.br>  
<http://www.netclock.com.br>  
<http://www.netclock.com.br>  
<http://www.netlens.com.br/>  
<http://www.netquadros.com.br>  
<http://www.netshoes.com.br>  
<http://www.nikkeycosmeticos.com.br>  
<http://www.ninjasomprofissional.com.br>  
<http://www.nonobier.com.br>  
<http://www.noteonline.com.br>  
<http://www.nsnsuplementos.com.br/>  
<http://www.oderco.com.br/>  
<http://www.oneshoes.com.br>  
<http://www.onlineeletronicos.com.br/>  
<http://www.onofre.com.br/>  
<http://www.painelshop.com.br/>  
<http://www.papelfantasia.com.br/>  
<http://www.passarela.com.br>

<http://www.passionfruitcompany.com.br>  
<http://www.paxdei.com.br>  
<http://www.pcinformatica.info/>  
<http://www.perfumagi.com.br/>  
<http://www.perfumariaseiki.com.br>  
<http://www.perfumes.com.br/>  
<http://www.petsupermarket.com.br/>  
<http://www.playtech.com.br>  
<http://www.pneusfacil.com.br/>  
<http://www.polishop.com.br>  
<http://www.poloar.com/>  
<http://www.pontodacasa.com.br/>  
<http://www.pontofrio.com.br/>  
<http://www.popipe.com.br/>  
<http://www.portcasa.com.br>  
<http://www.posthaus.com.br/>  
<http://www.poupafarma.com.br/>  
<http://www.prabebe.com.br>  
<http://www.prakaza.com.br>  
<http://www.premiershop.com.br>  
<http://www.puromalte.com.br>  
<http://www.quantastore.com.br/>  
<http://www.ramsons.com.br/>  
<http://www.reference.com.br/>  
<http://www.reidoterco.com.br>  
<http://www.relativa.com.br>  
<http://www.replicasitalianas.com.br>  
<http://www.resultdecorativos.com.br/>  
<http://www.ricardoeletro.com.br/>  
<http://www.rihappy.com.br/>  
<http://www.rochajoias.com.br>  
<http://www.rpaz.com.br/>  
<http://www.rmaquinas.com.br>  
<http://www.rsdiscus.com.br>  
<http://www.sabornatural.com.br/>  
<http://www.sacks.com.br>  
<http://www.sacrismalhas.com.br/>  
<http://www.sadeperfumes.com.br/>  
<http://www.sandrorosa.com.br>  
<http://www.santaefigeniashop.com.br>  
<http://www.santajerusalem.com/>  
<http://www.santaluciaonline.com.br>  
<http://www.saopaulomania.com.br>  
<http://www.sapatoversatil.com.br/>  
<http://www.sdmlentes.com.br/>  
<http://www.sebosonline.com/>  
<http://www.sensualsexshop.com.br/>  
<http://www.sepha.com.br>  
<http://www.sexshopcenter.com.br/>  
<http://www.sexshopfacil.com.br>  
<http://www.sexshoponline.com.br/>  
<http://www.sexshopousadia.com.br>  
<http://www.sextoy.com.br>

<http://www.shopfato.com.br>  
<http://www.shopluxo.com.br>  
<http://www.shopmania.com.br>  
<http://www.shoppwatch.com/>  
<http://www.shoptimao.com.br>  
<http://www.shuzu.com.br/>  
<http://www.siciliano.com.br/>  
<http://www.siciliano.com.br/>  
<http://www.sjoartigosreligiosos.com.br>  
<http://www.skinzilla.com.br>  
<http://www.smallgadget.com.br/>  
<http://www.smallgadget.com.br/>  
<http://www.socd.com.br>  
<http://www.sofutebolbrasil.com>  
<http://www.somlivre.com/>  
<http://www.specialbeer.com.br>  
<http://www.sportson.com.br/>  
<http://www.starfish.com.br>  
<http://www.stixx.com.br/>  
<http://www.strar.com.br/>  
<http://www.strommagazine.com.br>  
<http://www.studiolightlustres.com.br>  
<http://www.superdelicia.com.br/>  
<http://www.taco.com.br>  
<http://www.taqi.com.br>  
<http://www.tbperfumes.com.br>  
<http://www.tecnomidia.com.br/>  
<http://www.tedmusic.com.br/>  
<http://www.temestilo.com.br/>  
<http://www.tiguana.com.br>  
<http://www.timbres.com.br>  
<http://www.tkaesportes.com.br/>  
<http://www.todamulher.com.br>  
<http://www.topinternacional.com.br>  
<http://www.toqueverde.com.br/floresonline/>  
<http://www.toymania.com.br>  
<http://www.tozaki.com.br>  
<http://www.trapemix.com.br/>  
<http://www.travessa.com.br/>  
<http://www.tricae.com.br>  
<http://www.rousseau.com.br>  
<http://www.tuttitipi.com.br>  
<http://www.ultrafarma.com.br/>  
<http://www.ultramanequins.com.br>  
<http://www.universofantasia.net/>  
<http://www.valedasfestas.com.br>  
<http://www.valishop.com.br>  
<http://www.vamosfalar.com.br/>  
<http://www.varanda.com.br/>  
<http://www.vascoboutique.com.br>  
<http://www.viablushop.com.br>  
<http://www.viallure.com.br>  
<http://www.vidaecor.com.br>

<http://www.videolar.com>  
<http://www.vilabrinquedo.com.br>  
<http://www.vinhosnet.com.br>  
<http://www.vinhosweb.com.br>  
<http://www.visaoecia.com.br>  
<http://www.vivreshop.com.br>  
<http://www.w8tuning.com.br/>  
<http://www.watchstar.com.br>  
<http://www.waz.com.br>  
<http://www.weril.com.br/>  
<http://www.wishfashion.com.br>  
<http://www.worldcigars.com.br/>  
<http://www.worldwine.com.br/>  
<http://www.xickos.com.br/>  
<http://www.zekkalentes.com.br/>  
<http://www.zelo.com.br/>  
<http://www.zuba.com.br>  
<https://www.americanas.com.br>  
<https://www.bigsupermercados.com.br>  
<https://www.boacoisa.com.br>  
<https://www.cliqueexshop.com.br/>  
<https://www.gazin.com.br>  
<https://www.submarino.com.br>